Biodiversity Metrics for Use in the Ecosystem Approach to Oceans Management

T.J. Kenchington and E.L.R. Kenchington

Department of Fisheries & Oceans,
Maritimes Region,
Ocean and Ecosystem Sciences Division,
Bedford Institute of Oceanography,
PO Box 1006, Dartmouth,
Nova Scotia B2Y 4A2
CANADA

2013

Canadian Technical Report of Fisheries and Aquatic Sciences 3059

Fisheries and Oceans
Canada

Pêches et Océans
Canada

Canada

Canadian Technical Report of Fisheries and Aquatic Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715-924 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. Il n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques*.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 456 de cette série ont été publiés à titre de Rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de Rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de Rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Canadian Technical Report of
Fisheries and Aquatic Sciences 3059

2013

BIODIVERSITY METRICS FOR USE IN THE ECOSYSTEM APPROACH TO OCEANS MANAGEMENT

by

T.J. Kenchington and E.L.R. Kenchington

Department of Fisheries & Oceans,
Maritimes Region.
Ocean and Ecosystem Sciences Division,
Bedford Institute of Oceanography,
PO Box 1006, Dartmouth,
Nova Scotia B2Y 4A2
CANADA

© Her Majesty the Queen in Right of Canada, 2013.
Cat. No. Fs 97-6/3059E ISSN 0706-6457

Correct citation for this publication:

Kenchington, T.J. and Kenchington E.L.R. 2013. Biodiversity metrics for use in the ecosystem approach to oceans management. Can. Tech. Rep. Fish. Aquat. Sci. 3059: vi+188p.

TABLE OF CONTENTS

TABLE OF CONTENTS iii

ABSTRACT v

RÉSUMÉ v

1 INTRODUCTION 1
- 1.1 THE NEED FOR METRICS OF BIODIVERSITY 1
- 1.2 OBJECTIVES OF THIS REPORT 2
- 1.3 STRUCTURE OF THIS REPORT 6

2 WHAT TO MEASURE AND WHY? 7
- 2.1 THE DIVERSITIES OF "DIVERSITY" 8
  - *2.1.1 Ecological Diversities* 8
  - *2.1.2 Biological Diversities* 13
- 2.2 USES AND INTERPRETATION OF BIODIVERSITY MEASUREMENTS 20
- 2.3 THE MEANING AND RELEVANCE OF ECOLOGICAL DIVERSITY 22
- 2.4 BIODIVERSITY CHANGE IN THE SEA 30
  - *2.4.1 Extinctions, Invasions and Extirpations* 30
  - *2.4.2 Depletions and Disruptions* 34
  - *2.4.3 Loss of Genetic Diversity* 35
  - *2.4.4 Summary* 37
- 2.5 A SPECTRUM OF DIVERSITIES 37

3 HILL NUMBERS AND OTHER THEORY 40
- 3.1 "TRUE" DIVERSITY 40
  - *3.1.1 Hill Numbers Resurrected* 40
  - *3.1.2 True Diversity Debates* 43
  - *3.1.3 Application and Recommendations* 46
- 3.2 DIFFERENTIATION OR β DIVERSITY 48
- 3.3 GENETIC DIVERSITY 54
- 3.4 SUMMARY OF DIVERSITY THEORY 57

4 MEASURING ECOLOGICAL DIVERSITY 58
- 4.1 COMMUNITIES, TAXOCENES AND SURVEY PROTOCOLS 59
  - *4.1.1 Communities: Diversity of Where?* 59
  - *4.1.2 Taxocenes: Diversity of What?* 61
  - *4.1.3 Temporal Cycles: Diversity of When?* 65
  - *4.1.4 Survey Protocols: Diversity How?* 66
  - *4.1.5 Units of Measurement in Field or Laboratory* 75
  - *4.1.6 Summary* 76
- 4.2 RICHNESS 77
  - *4.2.1 Curvilinearity in Species Accumulation* 77
  - *4.2.2 Determining Sample Species Richness and Species Density* 84
  - *4.2.3 Community Species Richness and Species Density* 85
  - *4.2.4 Rarefaction* 89

4.2.5 Some Common Errors 92
4.2.6 Reporting Species Richness 96
4.3 ECOLOGICAL DIVERSITY *SENSU STRICTO* 96
4.3.1 Determining Sample Diversity 98
4.3.2 Determining Community Diversity 100
4.4 EVENNESS 103
4.5 EXPERIMENTAL DESIGN AND THE DRAWING OF INFERENCES 113

5 MEASURING OTHER FACETS OF BIODIVERSITY 113
5.1 SPECIES RELATIVE ABUNDANCES & COMMUNITY COMPOSITIONS 114
5.2 COMMUNITY AND SPECIES ABUNDANCES AND BIOMASSES 117
5.3 PHYLOGENETIC AND TAXONOMIC DIVERSITIES 120
5.4 FUNCTIONAL TRAIT DIVERSITY 129
5.5 BIOTIC INDICES 131
5.6 HABITAT AND ECOSYSTEM DIVERSITY 134
5.7 SIZE COMPOSITIONS AND SIZE SPECTRA 134
5.8 TROPHODYNAMICS 139
5.9 DEMOGRAPHICS 144
5.10 CONSERVATION STATUS 145
5.11 INTRA-SPECIFIC DIVERSITY 146
5.12 SUMMARY 148

6 TOWARDS QUANTIFICATION OF BIODIVERSITY 151

REFERENCES 154

APPENDIX I: GLOSSARY AND LIST OF SYMBOLS 175

APPENDIX II: DETERMINING SPECIES DIVERSITY FROM AN EXISTING DATABASE 184

ABSTRACT

Kenchington, T.J. and E.L.R. Kenchington. 2013. Biodiversity metrics for use in the ecosystem approach to oceans management. Can. Tech. Rep. Fish. Aquat. Sci. 3059: vi+188p.

The Convention on Biological Diversity drives a need for biodiversity measurement, often supposed to mean measurement using simple metrics. This report considers what should be measured and provides an overview of both recent theoretical developments and the available indices, including an examination of the metrics of richness, evenness and their combination – the "diversity" of quantitative ecology. We conclude that those three are required when monitoring and mapping biodiversity but only to meet non-specialist expectations. For that purpose, we recommend species richness, the Exponential Shannon diversity index and Heip's index of evenness. Otherwise, simple metrics cannot usefully capture the complexities of ecosystems. Hence, we recommend that biodiversity be quantified through assessments that use more advanced analytical tools. Application of simple metrics should be limited to illustrating conclusions reached through more powerful and more reliable approaches.

RÉSUMÉ

Kenchington, T.J. and E.L.R. Kenchington. 2013. Biodiversity metrics for use in the ecosystem approach to oceans management. Can. Tech. Rep. Fish. Aquat. Sci. 3059: vi+188p.

La Convention sur la diversité biologique souligne un besoin en matière de mesures de la biodiversité, qui signifient bien souvent des mesures effectuées à l'aide de paramètres simples. Ce rapport examine les mesures qui devraient être effectuées et offre un aperçu des récentes avancées théoriques et des indices disponibles. Il comprend également un examen des paramètres de la richesse, des paramètres de l'équitabilité et des paramètres de leur combinaison : la diversité de l'écologie quantitative. Le rapport conclut que ces trois paramètres sont nécessaires pour surveiller la biodiversité et en dresser une carte, mais seulement pour répondre aux attentes de non-spécialistes. À cet effet, le rapport recommande la richesse en espèces, l'exponentielle de l'indice de diversité de Shannon et l'indice d'équitabilité de Heip. Sans cela, les paramètres simples ne parviennent pas à saisir adéquatement la complexité des écosystèmes. C'est pourquoi le rapport recommande que la biodiversité soit quantifiée au moyen d'évaluations faisant appel à des outils analytiques plus sophistiqués. L'utilisation de paramètres simples ne devrait servir qu'à illustrer les conclusions tirées à l'aide d'approches plus efficaces et plus fiables.

1 INTRODUCTION

1.1 THE NEED FOR METRICS OF BIODIVERSITY

The Convention on Biological Diversity ("CBD"[1]) was opened for signatures at the United Nations Conference on Environment and Development in 1992 in Rio and entered into force the following year – Canada being among the first signatory parties. Over the two decades since, the conservation of biological diversity, or "biodiversity" as it is now commonly known, has become a major policy issue underlying a wide variety of decisions. Both scientific advice in support of management and the strategic research on which that advice is built have grappled with describing assorted kinds of biodiversity and understanding their roles in supporting valuable ecosystem goods and services. Simultaneously, the evident failures of fisheries management in the 1990s drove more attention to the long-standing recognition that single-species approaches are less than ideal. The proposed solution is the ecosystem approach to fisheries management, which has emerged gradually – FAO's *Code of Conduct for Responsible Fisheries* of 1995 and its Reykjavik *Conference on Responsible Fisheries in the Marine Ecosystem* in 2001 being key steps in the development. Means to fully implement the approach remain elusive, however, despite much on-going effort.

The need to demonstrate that nations have met the targets of the CBD creates a requirement for measurement of temporal change in biodiversity. Some authors have suggested that implementation of ecosystem approaches will likewise require measurable indicators (e.g. Cury *et al.* 2005). Meanwhile others have sought to map biodiversity in support of the design of MPA networks (e.g. Horsman *et al.* 2011). In the European Union, yet more stringent demands for quantification have been created by the Water Framework Directive, of 2000, and the Marine Strategy Framework Directive, of 2008. There has thus been a considerable, global effort seeking quantitative, operational definitions of biodiversity, to match the rather broad meaning espoused by the CBD, as well as both measurable metrics that can adequately represent the defined term and means to apply those metrics in practice. Further, parallel effort has been given to the measurement of the ecosystem effects of fishing – a related and overlapping, though not identical, topic. Much of that overall drive has been carried forward through national and regional initiatives (e.g. Royal Society 2003; Rees *et al.* 2008 and following papers; ICES 2009, 2010, 2011a, b, c, 2012b; Shin & Shannon 2010 and following papers; Hering *et al.* 2013 and following papers[2]) but there is also a global Biodiversity Indicators Partnership ("BIP"), which was established in 2006 to develop means for monitoring progress towards the CBD's 2010 targets and which now works in support of the

---

[1] Symbols and abbreviations used in this report are presented in Appendix I, which also summarizes our particular narrow usage of some terms that are elsewhere commonly given broader meanings.

[2] The special issue of *Hydrobiologia* introduced by Hering *et al.* (2013) was published too late to be fully considered during the preparation of the present report.

Convention's 2020 Aichi Biodiversity Targets and other multilateral agreements (BIP 2010, 2011). Meanwhile, development of new metrics of aquatic ecosystem condition continues apace (e.g. Borja *et al.* 2009) and quantitative ecologists have maintained their academic debates over the nature of diversity, along with appropriate ways to measure and compare the diversities of different ecosystems for research purposes. Those debates have now extended over half a century but, in wake of the CBD's advent, they have grown in scope without losing any of their old intensity (cf. Magurran & McGill 2011). Canadian scientists have been deeply involved in many aspects of this complex of initiatives and some of the outcomes have been applied by the Department of Fisheries and Oceans. DFO, however, has not yet adopted standard quantitative metrics of biodiversity, nor has its measurement been incorporated into the Department's routine data collection and reporting.

1.2 OBJECTIVES OF THIS REPORT

In this report, we aim to complement rather than supplement those prior efforts. The scientific literature on ecological and biological diversity is vast and we have not made any attempt to review it in its entirety, though we have sought to capture the major themes of recent research, while also returning to matters that were once the common knowledge of quantitative ecologists but which seem to have too often been forgotten in recent decades. Our aim is to inform and advise both those analysts within DFO who are charged with developing measures of biodiversity, in support of management decision-making, and those who must receive, understand and apply the resulting outputs. We have attempted to present the material in such a way that it will be comprehensible by each target audience. As will be explained in the Section 2, however, some three decades ago diversity measurement was largely abandoned by those conducting research into quantitative ecology, being left to practitioners and, subsequently, the lay public. That transition has led to much confusion and a serious loss of what was once professional understanding – a situation exacerbated by the deficient textbooks (e.g. Magurran 1988, 2004) published since leading ecologists turned their attention elsewhere. In consequence, the measurement of biodiversity is widely misunderstood, even within the scientific community, and much that is now commonly accepted must be set aside. Since users of this report have no reason to accept our recommendations and rebuttals at face value, we have deemed it necessary to explain the reasoning behind them, sometimes at length.

Our focus here is limited to "biological diversity" as that is defined in the CBD. Hence, while we have considered proposed means of monitoring the ecosystem effects of fisheries and other aspects of biological conservation, we make no attempt to address those fields in their entirety. We thus differ from the work of the BIP, which explicitly intended that its "biodiversity indicators" should cover more than just the measurement of biodiversity and should extend to monitoring of anthropogenic pressures and conservation actions (BIP 2011).

Within the topic of biodiversity, we confine ourselves to its measurement (meaning the process of developing quantitative values, more often through estimation than literal measurement) using "indices", by which we mean quantifiable metrics that summarize multiple measured variables (cf. ICES 2010). The well-known Shannon diversity index, *H'*, is a classic example, in that it reduces a set of species' relative abundances to a single scalar value. In general, we do not consider mere summary statistics to be indices, though some cannot be ignored (e.g. species richness, *S*, which can be nothing more than a count of the species present). Conversely, we do not accept the outputs of more complex analyses, such as multivariate ordinations, as being indices for our present purposes. We understand indices to be quantitative tools that could be applied to many data sets from various ecosystems, in contrast to what we term "indicators", which are situation-specific measures and may (but need not) invoke some index. This usage of "index" and "indicator" is widespread (e.g. Pomeroy *et al.* 2004; Rice & Rochet 2005; IOC 2006; Kenchington 2010; Kershner *et al.* 2011) but not universal (e.g. Backer 2008; Johnson 2008; Marques *et al.* 2009; BIP 2011). The "Large Fish Indicator" of Greenstreet *et al.* (2011), for example, is here considered an index, while its developers have applied it as what we would consider an indicator of aspects of the North Sea ecosystem. Furthermore, the IOC/UNESCO initiative *IndiSeas* (Shin & Shannon 2010) concerned itself with what it termed "indicators", some of which we consider indices while the rest are mere summary statistics (though the *IndiSeas* team did not firmly distinguish concept from application, blurring the line between index and indicator). Meanwhile, ICES (2010) took a rather different tack by which an "index" (given the same meaning as here) becomes an "indicator" if it responds to a particular anthropogenic pressure[3].

While we focus on the available indices, we do not consider them in isolation but rather in a context of the availability and quality of data to which they might be applied, as well as the ecological meaning that they bear and could bring to management decisions. Ours is thus a less abstract perspective than that of many academic reviews of the topic. In practical applications, meaning in attempts to develop specific indicators for some particular purpose, any index must be set within a complex of policy choices and stakeholder perspectives, as well as a context of multiple pragmatic issues concerning field-survey protocols, dataset management, monitoring budgets and more – all of which is at least as important as the selection of which metric to use. Frameworks for developing such contexts for biodiversity measurement have been discussed by the Royal Society (2003) and, extensively, by the BIP. Our focus, however, is necessarily more generic and we leave much for examination during future specific applications. Hence, all we offer here is one step in a lengthy process leading to biodiversity measurement – albeit a vital step.

---

[3] For the purposes of this report, we entirely set aside the established usage of fisheries science in which an "index" is any measure proportional to some variable of interest, such as the mean catch per set in a survey being used as an "index" of abundance or biomass.

Furthermore, this report is primarily concerned with Canadian aquatic environments, though we have drawn on ideas developed elsewhere and for other types of ecosystems. At an abstract level of reviews of indices, there are parallels between the conservation biology of isolated glacial lakes and that of the terrestrial systems for which much of the theory of biodiversity conservation was developed. The ecosystems of the open sea, along with the principal anthropogenic pressures that act on them, are radically different – qualitatively as well as quantitatively. We therefore devote space to specifically marine applications of the indices, not because those are more important to the Department but because they require more development here.

In the debates surrounding implementation of both the ecosystem approach to fisheries management and the CBD, much attention has been given to means for selecting a suite of indices to employ in routine monitoring. On its face, that appears to be a critical step, given the great number of alternative metrics that have been proposed over the decades. Certainly, the choices of which facet of biodiversity to measure and which index to use in its measurement will influence the results obtained, and the selections must therefore be made with care, then fully justified and validated (Heink & Kowarik 2010). Those tasks extend far beyond our current focus, requiring tripartite cooperation between scientists, policy-makers and stakeholders (Rice & Rochet 2005; Heink & Kowarik 2010), but the recommendations that we offer here must be relevant to the broader process.

While conservation biologists have tended to follow other routes (see partial review by Heink & Kowarik 2010), the selection approach most often invoked amongst fisheries scientists centres on a detailed framework process advanced by Rice and Rochet (2005) that was constructed around nine proposed criteria: concreteness, theoretical basis, public awareness, cost, measurement, historical data, sensitivity, responsiveness and specificity. While that framework does not appear to have often been applied in full, its criteria do serve as a useful classification of most of the issues to be considered when selecting among alternative metrics – though others were not emphasized as much as they might have been by Rice and Rochet (2005). For example, Rice (2000) had demanded an understanding of a metric's sensitivity to the ecosystem properties that it summarizes and confidence that those properties are the relevant ones. Likewise, Heink and Kowarik (2010) emphasized "relevance". As will be explained below, biodiversity as such cannot be measured, in the sense of being reduced to a single number. Only particular facets can sometimes be represented in numbers – and only for selected portions of a biotic community. It follows that two aspects of "relevance" must be addressed: those facets of biodiversity that are relevant to the issue at hand must be determined and their selection justified, but the relevance of some index or indicator, as a measure of one of the facets deemed relevant, must also be established – the latter generally requiring empirical confirmation within each specific setting (cf. Heink & Kowarik 2010). In addition, Rice (2000) called for objective determination of whether differences in observed values are meaningful, along with a way to partition causality among various forcing factors, while Heink and Kowarik (2010) saw a need for what they termed "validity" or "the quantitative connection between the evidence and the interpretation of the results".

Our work must fall within these various *desiderata* and we have taken into consideration the combination of criteria recommended by Rice (2000), Rice and Rochet (2005) and Heink and Kowarik (2010). It is not, however, our task to select a suite of indicators for any one purpose nor even a generic set of indices from which indicators should be developed. Rather, our aim is to provide scientific recommendations to inform future selection processes. Thus, while it would be futile to consider the available indices without regard to their application and in ignorance of the concerns of managers and stakeholders, our focus is on scientific aspects alone, leaving other considerations to a later phase of decision-making. Moreover, despite the very large number of indices of biodiversity that have been suggested over the past half-century, few (if any) meet the standards demanded by Rice (2000) and Heink and Kowarik (2010). Thus, in contrast to the attention on means for selecting among many alternative metrics that is frequent in the literature, we see a challenge in finding any useful indices at all. That scarcity of indices able to meet Rice's (2000) standards largely negates the need for Rice and Rochet's (2005) selection framework, allowing us to offer recommendations which should prove useful, despite our not using the methodology advanced by those authors.

In practical applications of biodiversity measurement, the criteria of cost and of the availability of historical data (cf. Rice & Rochet 2005) will often over-ride most other considerations – budget constraint being a particular issue when monitoring offshore marine ecosystems because of the very high cost of research ships. While we have not allowed those two criteria to overwhelm our work, we have given some special emphasis to the use of groundfish trawl survey data in measuring biodiversity, that being the most temporally and spatially extensive data type available for the majority of the aquatic area under Canadian jurisdiction. Moreover, the various groundfish survey series are on-going and will continue for the foreseeable future, meaning that their data streams will be available for biodiversity monitoring at minimal incremental cost.

Since our treatment is generic, it tends towards the abstract and theoretical, while we have sought approaches that can be applied generally and not only in specific cases. We have therefore unapologetically dismissed metrics that, from an examination of their mathematical expressions, should not effectively and efficiently quantify some facet of biodiversity. That may be a valuable counter to the too-common practice of promoting some index on the grounds that it works adequately with one particular data set. Nevertheless, theoretical imperfections in a metric do not mean that it would necessarily return misleading results if applied to some particular data series. There is scope for case-specific applications of some indices but they require individual exploration of the scale of their errors from a research orientation (perhaps as part of the biodiversity assessments proposed in Section 6), rather than immediate routine implementation. We have not attempted such explorations here.

1.3 STRUCTURE OF THIS REPORT

In both textbooks and primary papers, the measurement of biodiversity is typically approached as a series of individual issues. Yet it is then all to easy to arrive at erroneous conclusions. Indices, the means to determine their values and the relevance of those values to policy questions must be viewed as a totality. Presentation of the issues from that perspective is, however, complicated by the topics forming what might (metaphorically) be seen as a "spherical network" of knowledge. There is no point in that sphere where an explanation can break into the structure of information without requiring the reader to have prior knowledge of other points. In this report, we resort to multiple internal forward or backward references among sections but it is nevertheless impossible for us to avoid all repetition. Some topics must be introduced briefly before others can be discussed, the former being elaborated thereafter, once the reader is equipped with an understanding of the latter.

Before engaging with the available indices of biodiversity, it is first necessary to consider the nature of what is to be measured. Thus, in Section 2 we examine the broad range of ideas that have been grouped under the "biodiversity" label – in which treatment we follow a largely chronological sequence. We find a spectrum of ideas, extending from the narrowly-defined concepts of quantitative ecology to a breadth of notions within which any topic of relevance to biological conservation can be accommodated under the rubric of "biodiversity". Section 2 also includes consideration of the uses to which measurements of biodiversity may be put, the ecological meaning that can be borne by those measurements, and the nature of spatio-temporal variations in the biodiversity of the open sea – contrasted as they are with those of terrestrial, freshwater and coastal systems.

In Section 3, we examine certain issues in biodiversity theory in greater depth, with a particular focus on recent developments in the scientific literature. Jost's (2006, 2007, 2008, 2010a, b) renewal of emphasis on the Hill Numbers is a particular focus, since that has greatly simplified choices amongst the indices developed by quantitative ecologists. We also dwell on the topic of β or differentiation diversity. That will rarely be measured when monitoring or mapping in support of management but an understanding of the topic is essential to comprehending the relationships between diversity and spatial scale – and hence in application of measurements of the former. In Section 3, we also consider genetic diversity, the concepts and indices of which have been developed by population geneticists somewhat independently of the work of quantitative ecologists, though both topics are now subsumed under the biodiversity of the CBD.

This report then turns, in Section 4, to the practical challenges of measuring richness, evenness and the "diversity" of quantitative ecology, with special attention to (temporal) trend monitoring and (spatial) mapping. That leads to specific recommendations of which indices and approaches should be used in the Department's monitoring and reporting. We provide both specific guidance on how the indices should be quantified and an explanation of those steps. In Appendix II, we offer directions for the application of our recommendations to existing datasets. Also in Section 4, we include a cautionary note about the limits to our advice, which covers only measurement, not the experimental designs necessary if comparable measurements are to be obtained.

Section 5 then offers much briefer accounts of the indices that have been suggested for measurement of other facets of biodiversity. That leads us to an overall conclusion that is not entirely novel and in retrospect seems unsurprising, though it was unexpected by us and will likely seem so to many readers. It is that even single facets of anything as complex as an ecosystem can rarely be usefully captured in single numbers. Hence, the indices that are the subject of this report have limited value in biodiversity measurement – meaning that much of the material assembled here is less a review of how to measure diversity and more an explanation of why it should not be measured using the commonly accepted approach of simple metrics. In our final summary, Section 6, we therefore offer an outline of the more-advanced assessments that are needed if biodiversity is to be quantified. We do not develop that proposal, which extends far beyond our present topic.

2 WHAT TO MEASURE AND WHY?

Before attempting to measure biodiversity, it is essential to understand what it is that should be measured and why the measurements are needed. Those issues are addressed in this Section 2, beginning with an overview of the multifarious concepts of diversity. That is followed by consideration of the purposes and value of measuring biodiversity, which leads into questions of the meaning and relevance of ecological diversity. Finally, we consider the nature of biodiversity change in aquatic ecosystems and particularly in the open sea, where it is fundamentally different from expectations built on terrestrial experience – with major implications for measurement.

This section focuses on what diversity is, as distinct from how it should be quantified. The latter is the topic of the remainder of this report.

2.1 THE DIVERSITIES OF "DIVERSITY"

In common English, the word "diversity" relates to "diverse" and means a condition of being different or having different kinds. It is a synonym of, among others, "unlikeness", "heterogeneity", "dissimilarity", "divergence" and "varied". Through the past three quarters of a century, those general senses have been taken up, though very much narrowed, in a complex of concepts much invoked in the ecological sciences and, more recently, in national and international policy goals. "Diversity", in its technical senses, has however proven to be one of those elusive ideas which seems simple at first sight but reveals unexpected complexities and shifts in its nature as it is examined more closely. Too often, those complications have been allowed to generate confusion. Indeed, within fifteen years of MacArthur's (1957) and Margalef's (1958) seminal papers on the topic, Hurlbert (1971) could label species diversity a "nonconcept" (though he offered neither evidence nor argument in support of that postulate), while the four decades since have seen the problems expand, particularly as scientific meanings of "diversity" have become entangled with the policy objectives of the lay public. So convoluted have matters become that Swingland (2001) insisted that "biodiversity" is undefinable, while Mayer (2006) cited multiple criticisms of the concept and suggested that it is best understood as something context dependent. That, however, furthers the risks of serious mistakes: as Leinster and Cobbold (2012) wrote: ""Politicians may understand diversity to mean one thing; the scientists advising them may use it to mean another. Misguided policies may be the result" (cf. Rodda 1993).

2.1.1 Ecological Diversities

The term "diversity" appears to have first entered the language of the life sciences with the work of Williams (in Fisher *et al.* 1943). At the time, early quantitative ecologists were struggling with understanding the differing numbers of species taken in samples of animals or plants, along with the varied abundances of individual specimens of each of those species within the sample. Fisher (in Fisher *et al.* 1943) developed the log series distribution to fit distributions of the numbers of species represented by n specimens in a sample (which have since become known as "Species Abundance Distributions" or "SADs"). In applying the new model, Williams (in Fisher *et al.* 1943) suggested that its α parameter should be called the "index of diversity" since it represented the increase in the number of observed species achieved by increasing the sample size by e. Unfortunately, that proposal depended on an assumption of the log series, while further experience showed that many sets of species abundance data do not well fit that, nor any other, particular distribution (cf. Hurlbert 1971; Magurran 1988, 2004). Besides, in many research settings it is the deviations from theoretical ideals which are most interesting.

Through the middle decades of the 20th Century, quantitative ecologists worked to lift ecology from mere natural history into a science with laws that could be expressed in the language of mathematics, following the precedents set by the physical sciences (Pielou 1975). To do so, they sought to reduce the almost-unlimited complexities of whole ecosystems to a few metrics that would allow comparisons among ecosystems and hence the construction of unifying theories. From MacArthur (1957) and Margalef (1958) onwards, alternative indices of diversity took a prominent place in that effort. "Species richness" (*S*), the number of observed species in a sample or a community, has been one favoured metric since the 1960s at least (Washington 1984) but it was seen to be an inadequate measure, if used alone, since a monoculture with a dozen other species so rare as to be barely detectable is clearly less diverse, in the broad sense of the word, than a community with fewer species in total but all of them reasonably abundant. Besides, richness gives equal weight to each species while most communities contain far more rarities than abundant species, hence richness is primarily a count of the number of rarities present (Magurran & Henderson 2011). Yet it is the few abundant species that are responsible for the bulk of ecosystem function, including production, energy fluxes and the like. Thus, richness was combined with a new concept, termed "evenness", "equitability" or (when inverted) "dominance" that reflects the relative abundances of the species present – typically a few being highly abundant, many being comparatively rare and some falling in between. Representing distributions of the relative abundances of species by two parameters rather than one was, however, intellectually dissatisfying and hence, as early as Margalef's work of 1951, the combination of richness and evenness was defined as "diversity" (Pielou 1974, 1977; Washington 1984). Many indices have been proposed as metrics of that diversity but much the most commonly used has been Shannon's:

$$H' = -\sum p_i \log_b p_i$$

where p_i is the relative abundance of the i^{th} of *S* observed species and *b* is any preferred base for the logarithm (usually either 2, *e* or 10). We return to that index later in this report but it can stand here as an illustration of the kinds of metrics used.

This "diversity" of quantitative ecology has sometimes been called "ecological diversity" (e.g. Pielou 1975; Washington 1984) and that term will be used here to distinguish it from the later "biological diversity" of the CBD – though we recognize that our usage is not universally accepted. Its original meaning, as a combination of richness and evenness, has not been entirely forgotten (e.g. Spellerberg & Fedor 2003; Tuomisto 2010a) but neither has that been maintained consistently, even within professional circles. As early as the 1970s, Hurlbert (1971) and Peet (1974) conflated richness, evenness, dominance and the diversity of Margalef (1958) or Pielou (1974, 1975, 1977) – all four of which Magurran (1988, 2004), in her influential textbooks, regarded as different aspects of a wider "diversity". More recently, she has described indices of richness or evenness as "'Classical' diversity measures" (Magurran 2011), even though they are not measures of ecological diversity, in its original sense, at all. She acknowledged that "there are many measures that blend richness and evenness" (Magurran 2011), without noting that only those measures are, strictly speaking, "diversity" indices, and went on to state that "Species richness is the iconic diversity measure" (Magurran and Henderson 2011; cf. Magurran 2004). Indeed, too many scientists appear to think that species richness is biodiversity, not simply an imperfect measure of one facet of it (e.g. Heink & Kowarik 2010). Rosenzweig (1995) went so far as to write an entire textbook about "species diversity" in which he addressed only species richness, rejecting indices that combine richness with equitability[4]. More recently, Maurer and McGill (2011) noted the idea that diversity is a function of richness and evenness, describing that (original and formally defined) meaning as a "traditional" interpretation or one "often intended". To add to the confusion, as is explained below (Section 3.1), richness is embedded within Jost's (2006) framework of diversity measures that has been, collectively, termed "true diversity", yet Tuomisto (2010a) has simultaneously advocated for Jost's (2006) vision while stoutly maintaining the original distinction between diversity and richness.

Such lax use of terminology does not aid understanding but neither would pedantic insistence on original definitions. Hence, in this report we use "ecological diversity" as a general label, broadly encapsulating all of Hurlbert's (1971) "nonconcept". When necessary, we distinguish the original meaning of the term, as it was understood by Pielou (1974, 1975, 1977), as "ecological diversity *sensu stricto*", its combination with richness and diversity then being "ecological diversity *sensu lato*".

Ecological diversity *sensu stricto* has been criticized for being responsive to variations in two different features of Species Abundance Distributions, both their numbers of species and the relative abundances of those species (e.g. Rosenzweig 1995; Magurran 2004; Dornelas *et al.* 2011). Richness and evenness can sometimes vary in opposing directions, maintaining near-constant ecological diversity *sensu stricto* despite changes in a monitored community, while somewhat less of the expensively-collected information in field data on species abundances is lost if SADs are summarized by two indices rather than one (Dornelas *et al.* 2011). There was, however, merit in the original (if overly ambitious) intent of reducing a SAD to a single metric and, with modern computational power, it is convenient to tabulate all three of richness, evenness and ecological diversity *sensu stricto*, then look for patterns of agreement or disagreement between them when they are applied across communities or across time within one community. Hence, we recommend retention of ecological diversity *sensu stricto* but also that, whenever possible, it be accompanied by measures of richness and evenness.

The concepts of ecological diversity were originally intended specifically for the summary of data on species abundances. When so used, they can be termed

---

[4] In a footnote justifying his position, Rosenzweig (1995) claimed that the rejected indices are used to make diversity a continuous variable and that the term "richness" was introduced by Hurlbert (1971), neither of which contentions has any discernable basis in fact, while he demanded that richness be known as "species diversity" – a most unfortunate proposal.

"species diversities". It appears not to have been long, however, before thought was given to measuring diversities of other units, such as higher taxonomic levels (e.g. genera), smaller units (e.g. intra-specific populations or alleles), groups of organisms classified by ecological traits or age, and even habitats (Magurran 1988). Indeed, MacArthur considered the diversity of trophic pathways as early as the mid-1950s (Washington 1984). So long as individual units (e.g. individual habitats) can be uniquely assigned to one of a set of discrete groups, the relative abundances of the members of those groups can be summarized using the same indices that are used as metrics of species diversity. There have been a number of studies which have pursued such approaches, though the overwhelming majority of applications of the indices of ecological diversity has been to species diversity. Population geneticists have made parallel calculations concerning allele frequencies but have evolved their own metrics for that purpose.

Through the 1960s, the concept of ecological diversity allowed quantification of such obvious trends as the greater variety of species at lower latitudes and the lesser numbers of them in harsh environments. The original focus was on seeking explanations for differences amongst ecosystems, once those had been summarized into diversity and a few other metrics. There were speculations, supported by reasoning which (with the benefit of hindsight) now looks remarkably naïve, that any of more complex habitats, longer evolutionary histories, higher productivity and greater environmental stability might lead to the larger diversities observed in some communities. Habitat complexity should increase the variety of niches and longer histories should allow species to adapt to occupy those niches. Higher overall productivity of an ecosystem should allow species to remain viable while being specialized to utilize one particular component of that system as an energy source, while stability should ensure that the particular source is always available (cf. Pielou 1974). Through the decades up to about 1980, some of those ideas were developed into hypotheses which were then tested using field data, leading to rejection of the naivety and a much deeper understanding of the processes that shape the richness and evenness of communities (cf. Pielou 1975; Rosenzweig 1995). However, the hoped-for overarching, quantitative theories of ecology remained elusive – even explanations for the major latitudinal gradient in ecological diversity remaining disputed to this day[5].

In parallel with those early explanatory efforts, other ecologists pursued the reverse ideas (cf. Washington 1984; McCann 2000). Instead of complexity, productivity and stability being seen as drivers of ecological diversity, it was suggested that: (1) higher diversity should lead to greater trophic efficiencies, through the presence of more-specialized producers and consumers, which would enhance both primary production and the transfer of energy to higher trophic levels, thus increasing production throughout the trophic pyramid, (2) higher diversity and hence ecosystem complexity should lead to greater community, and perhaps even ecosystem, stability and (3) anthropogenic disturbance, or perhaps a subset of particular disturbances, should depress diversity, such that an index of diversity could be used as a measure of disturbance. Those suggestions were largely based on theoretical reasoning, always a questionable guide when dealing with complex systems, while ecosystems and the factors which shape them have yet to be successfully reduced to mathematical expressions, rendering theorizing especially speculative. Empirical support for those supposed effects of greater ecological diversity was thin, as it remains even today (see Section 2.3).

---

[5] Though Rosenzweig (1995) argued cogently that much of that gradient results from the greater surface area of the planet in low-latitude climate zones

The 1960s and 1970s were, as Rosenzweig (1995) put it, the "glory days" of research into ecological diversity and proved to be an exciting era for quantitative ecology, with much new understanding of ecosystems resulting. It is fortunate that two extended summaries of the fruits of that work were published, one contemporary (Pielou 1975) and the other a retrospective (Rosenzweig 1995). Nevertheless, by the 1980s the search for fundamental laws of ecology could be seen to be largely sterile, while ecological diversity had proven to be a rather uninformative measure. Quantitative ecology therefore moved on. By the early 1990s, Rosenzweig (1995) could jest that questions of species richness had become a "dinosaur" while, when Magurran (1988) tried to answer her own question "Why diversity?", she could only offer that it was a central theme in ecology, which was circular reasoning, that it was a matter of much debate, which was only promotion of fruitless argument, and that measures of diversity "are frequently seen as indicators of the wellbeing of ecological systems". That they were (and are) seen as such is certain, though their efficiency and effectiveness in that role are dubious in both terrestrial and freshwater systems. In open-sea ecosystems, it is clear that they are neither. When she came to state the practical value of diversity measurements, Magurran (1988) gave no hint of the original goals of the quantitative ecologists but offered indices of ecological diversity only for monitoring the effects of certain anthropogenic stressors (primarily chemical pollutants and eutrophication) and as one (but not the only) axis on which to score different alternatives in protected-area planning. Even then, she admitted that moderate eutrophication of nutrient-deficient systems increases ecological diversity, while she argued that it is the diversity of habitats, not species diversity, which is the best guide to determining where conservation efforts should be focused (Magurran 1988).

Rather than pursuing such a narrow focus, many other research ecologists of that decade simply moved away from the use of ecological diversity. Around 1980, increasing computing power and the availability of packaged statistical software opened the option of far more powerful ways of examining and summarizing the complexities of ecosystems – a process subsequently carried much further still with the advent of geostatistical approaches. Even the task of reducing species abundance data to comprehensible formats was better done through ordination than *via* the calculation of scalar values of richness, evenness and diversity. Those measures thus fell out of use in cutting-edge quantitative ecology. It is symptomatic that Pielou, in the mid-1970s, emphasized diversity indices in her undergraduate and graduate textbooks of quantitative ecology (Pielou 1974, 1977), while writing her manual on ecological diversity (Pielou 1975), but in 1984 produced a primer on multivariate ordination and classification (Pielou 1984).

It is also symptomatic of Rosenzweig's (1995) "glory days" that he and Pielou (1975) could each summarize the breadth and depth of ecological understanding that emerged from the 1960s and 1970s without mention of anthropogenic declines in diversity, which were to become the dominant theme by the 1990s – and that despite Rosenzweig (1995) repeatedly stressing his personal concern about the rate of loss. While the environmental movement had become interested in diversity questions in the 1960s, as late as 1980 ecologists could address the science of undegraded systems, as Pielou (1975) and Rosenzweig (1995) did, rather than focusing on anthropogenic degradation.

2.1.2 Biological Diversities

Although thus largely discarded for research purposes 30 years ago, an outline of ecological diversity continued to be taught to students and, regardless of the limitations of the available metrics, their values have been calculated (or far more often miscalculated) and reported whenever appropriate data are gathered. Richness and diversity values have been presented in innumerable field reports, where they serve to reduce a species list to a single number, in environmental assessments as a way to rate the areas under consideration for anthropogenic development, and in conservation planning, when selecting areas that may merit enhanced protection. From that wide, if shallow, familiarity amongst practising scientists, some uncertain understanding of the notion of ecological diversity spread out to the interested lay public, particularly members of the environmental movement. There, the dubious suggestions that ecological diversity is monotonically related to stability and production, but especially that it could be related inversely to human impacts, took root and expanded in new directions. As is to be expected of ideas developed outside the research community, the result has been a diffuse complex of vague and ill-formed notions that defy summary. Although the label "diversity" was applied by ecologists to so many different ideas in its first decades of use that Hurlbert (1971) could call it a "nonconcept", what has grown from it is immeasurably more confused and only some of its themes can be sketched in here. Importantly, despite subsequent shaping by conservation biologists and their introduction of a degree of rigor, these new ideas remain rooted in the worlds of policy makers, environmental advocates and the general public. They must be understood in that context and not solely in a scientific one – "the jargon word 'biodiversity' is, by its very origin, fundamentally indefinable, being a populist word invented for convenience" (Swingland 2001).

As it diffused out from the scientific community, vague public comprehension of ecological diversity became merged with two other, far older, drivers. One is that variety, in contrast to plainness, appeals to human aesthetic senses – though individuals can be selective about where and when variety or plainness is preferred. Visual variety within ecosystems is directly related to, could even be quantified through, ecological diversity. Thus, diversity can be equated to beauty, at least in the eyes of some beholders. Secondly, many citizens of wealthy, post-industrial nations desire "wildness" and "naturalness" (the two being closely related in terrestrial environments, though not in the sea), even if only when they are maintained at a safe distance from an individual's home and economic interests. That desire for "wildness" became a major political force through the later 20th Century. In North America, it is commonly traced to the writings of Henry David Thoreau and specifically his 1854 book, *Walden*. However, a similar preference for a contrived "naturalness", among those who could afford it, can still be seen in the English landscape architecture of the mid-18th Century. With more emphasis on "wildness", it is also prominent in artworks from the Romantic Era of a few decades later. The origins of this attractiveness of the wild have been debated but seem to include a reaction to expanding industrialization and urbanization, many people responding to the growing human ability to control nature by seeking to protect such "wildness" as remains. During the 1970s and 1980s, the common (though erroneous) supposition that ecological diversity is inversely and monotonically related to the degree of human impact on ecosystems led to the idea that it is also directly related to "wildness", giving diversity the appearance of a measurable and scientifically-based surrogate for that policy goal. So strong did that appearance become that at least one proponent of "wildness" had to warn that diversity is an inadequate measure and that the degree of "wildness" must be judged directly, in terms of the absence of human interventions (Turner 1994).

The hypotheses which postulate that productivity and stability of ecosystems are driven by ecological diversity, were speculative but were too readily presented to the public as though they were proven conclusions. Those ideas could be combined with the mistaken notion that anthropogenic impacts in general (not just particular degrees of certain impacts) reduce diversity, leading to a supposition that human activities reduce both productivity and stability of ecosystems through the medium of diversity loss. It was then but a small step to equate higher diversity with a "good" state of nature, as well as a "wilder" one, sometimes to the extent of simply disregarding the natural differences in the diversities of various communities. From there, an understandable but fallacious transformation led to a supposition that the "goodness" of an ecosystem can be measured by, perhaps that it actually is, the diversity of the communities in that ecosystem. A converse reversal of logic has since made anything which is "good" about the state of an ecosystem into an aspect of "diversity", even if such factors have no discernable relationship to the relative abundances that are the sole foundation of ecological diversity.

Through the 1970s, this dimly understood and poorly classified mix of political, scientific and aesthetic considerations overlying the diversity of the quantitative ecologists came to be labelled "natural diversity" or, more often, "biological diversity"[6]. The intent was perhaps only to distinguish the diversity of biological systems from the many other kinds of diversity encountered in everyday life (e.g. diversities of goods available in stores and of socio-economic status) but the new term serves also to distinguish the broad, public understanding of biotic variability from the narrow and (relatively) rigorous concept of ecological diversity. While many authors would choose to employ the terms differently, in this report "ecological diversity" is reserved for that latter concept while the broad one is denoted "biological diversity".

That term, along with some of the breadth of public perception associated with it, returned to academic use during the 1980s through the then-emerging discipline of conservation biology. It was introduced to the scientific literature in that context by one of the fathers of the field, Thomas E. Lovejoy of the World Wildlife Fund, in the early textbook *Conservation Biology* (Soulé & Wilcox 1980). The now-familiar contraction, "biodiversity", is said to have been coined in 1985 though it first appeared in print in 1988 (Wilson 1988). The nations of the world began signing the Convention on Biological Diversity less than four years later.

Unfortunately, although conservation biologists have been more rigorous than the general public in their use of "biodiversity", the discipline has not agreed on a definition of the term. While some have sought to make it into a broad umbrella, others have stuck closer to the concept of ecological diversity. The one definition which has achieved some general currency is that adopted in the CBD:

> Biological diversity means the variability among living organisms from all sources including *inter alia*, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems.

Subsequent published, technical definitions tend to be similar, the Canadian Biodiversity Strategy repeating the CBD's verbatim.

Thus defined, biological diversity might differ little from ecological diversity, arguably only by demanding explicit attention to diversities at scales both larger and smaller than species – a requirement long acknowledged by quantitative ecologists, though too rarely considered in practical applications. It is perhaps important that "within species" diversity is mentioned, since that draws genetic diversity into the same fold as diversity at species and higher levels. Perhaps the

---

[6] Etymologically, "biological diversity" is often traced to R.F. Dasmann, whose 1968 book *A Different Kind of Country* was an example of the then-common genre of extreme environmentalist over-reaction to the damage caused by unbridled industrial development in the 1950s and 1960s. Magurran (2004), however, has traced "biological diversity" back to a 1955 paper in *Systematic Zoology* by N.L. Gerbilskii and A. Petrunkevitch. She identified other papers published between 1955 and 1968 which also used the term.

CBD definition could also be seen as incorporating the oft-forgotten diversity of biotic processes, such as gene flow, interspecific interactions, natural disturbances or nutrient cycles – functional aspects of ecosystems that are at least as important to conservation as are the compositional aspects. Noss (1990) objected to a precursor of the CBD definition on the grounds that it did not encompass variability in processes and that deficiency was not overtly addressed in the final wording of the Convention but the topic might be seen as implicitly included.

It seems unlikely, however, that opinion leaders in the environmental movement would be satisfied with such a narrow interpretation of the CBD. In ecological diversity, all species are replaceable, provided that richness and evenness are maintained. The metrics of richness, evenness and diversity not only disregard the names and identities of the species in the community under study but those are not even given unique identifiers within the calculations. Hence, such an environmentalist *cause célèbre* as the replacement of the California sardine with anchovy (cf. MacCall 2011) need not have affected the richness, evenness or ecological diversity *sensu stricto* of the California Current ecosystem. Likewise, while the final extirpation from an ecosystem of a very rare species necessarily marks a loss of richness, it is unlikely to cause a noticeable change in measured ecological diversity *sensu stricto* since extreme rarities have very small relative abundances, if they are even recorded in surveys at all. The last right whale could disappear from the North Atlantic forever without detectable change in that Ocean's cetacean diversity. It is true that the loss of the last individual of a species cannot have much ecological significance, since no solitary organism can play a significant role in any but the smallest of ecosystems, and hence it is no criticism to note that ecological diversity is unresponsive to such losses. Extinctions and extirpations nevertheless remain matters of grave concern for environmental policies and managers, as much as for environmentalists. Hence, in marked contrast to ecological diversity, public understanding of biodiversity is very attentive to the identities of the species present and to the persistence of the rarities. Equally, evenness and ecological diversity *sensu stricto* consider only relative abundances, while richness is entirely blind to abundances of any kind. None of the three would respond as the overall abundance of the organisms in a community declined towards zero, provided that the relative status of the species was maintained. In contrast, most members of the interested public would be likely to see such a decline as a calamitous loss of biodiversity.

Some of those wider perspectives could perhaps be accommodated within more-liberal interpretations of the CBD definition but, even if others cannot, neither lay nor professional usage has been much constrained. No ecologist would mistake the "goodness" of higher diversity with diversity being "goodness", nor seek to quantify the visual appeal of a landscape, but with public policy tending to treat the term "diversity" as a synonym for a general "goodness" of an ecosystem, some authors have placed every facet of an ecosystem that is important to an ecological or conservation goal under the "diversity" umbrella. Thus, in its consideration of biodiversity measurement, the Royal Society (2003) mentioned species richness (as a "surrogate for overall biodiversity") and genetic variation but also rarity, extinction risk, abundance of selected species, spatial extent of habitats and ecosystem processes and functions – none of the latter involving measurement of variability. When introducing their new textbook, Magurran and McGill (2011) explicitly extended the measurement of what they termed "biodiversity" to include occurrence, range, vulnerability, functional traits and phylogenetic diversity, not merely the richness, evenness and distribution of species abundances. Likewise, Chown and McGeoch (2011) mentioned richness but also turnover, abundance, range size and body size in the context of "biodiversity" measurement. Stepping even further from the vision of the early quantitative ecologists, Dornelas *et al.* (2011) addressed the topic of "biodiversity and disturbance" while deprecating the use of indices of ecological diversity *sensu stricto*. They claimed "decades of admonitions against the use of these metrics", which seems a biased reading of the literature but of immediate interest simply because Dornelas *et al.* (2011) could suppose it to be true.

In the management arena, a consortium of government agencies in the United Kingdom has begun producing an annual series of booklets under the title *UK Biodiversity Indicators in Your Pocket*. The 2012 version included time series of species richnesses of plants in various environments, indicators that are clearly related to ecological diversity, but also such measures as the abundances of various groups of birds, bats and butterflies, the proportion of fish above 40 cm length in one survey series, the length of rivers deemed of good biological quality, the timing of the arrival of spring, the percentage of woodland area certified as sustainably managed, the percentage of identified sensitive areas that exceed critical acidification or eutrophication loads, an index of volunteer time spent in conservation organizations, and the amount of public expenditure on biodiversity (Anon. 2012). Similarly, the United Nations Environment Program has recommended that indicators of biodiversity include species abundances but also their distributions, the extent of protected areas, the status of threatened species, alien invasions, fragmentation of ecosystems, and the well-being of humans in biodiversity-dependent communities (ICES 2010). For monitoring the achievement of CBD biodiversity goals the BIP recommended "headline indicators" that spanned trends in abundances, distributions, ecosystems and habitats, coverage of protected areas, status of threatened species, land and water areas under sustainable management, ecological footprints, trends in alien invasives, the Marine Trophic Index, water quality, trophic integrity, well-being of human communities, status of (human) linguistic diversity and indigenous knowledge, benefit sharing, technology transfer and more. The CBD definition notwithstanding, variability of non-human living systems was only invoked in two "headline indicators": "Trends in genetic diversity of domesticated animals, cultivated plants, and fish species of major socioeconomic importance" and "Biodiversity for food and medicine" (BIP 2010). While all of the other indicators have understandable relevance to conservation policies and programs, the relationships of most of them to the diversity of life, or to the formal definition of biological diversity, are indirect at best[7].

The CBD process itself has adopted some of these wider perspectives, thus interpreting "biodiversity" with a rather broader meaning than seems intended by the wording of the Convention's definition. Even the original CBD text called for signatories to protect and where necessary restore ecosystems, maintain viable populations and promote recovery of threatened species. The subsequent *Strategic Plan 2011–2020* incorporates five "Goals" and twenty "Aichi Biodiversity Targets". While four of the former are framed in terms of biodiversity, the "Targets" include mention of, *inter alia*, pollution, invasive alien species, rate of loss of natural habitats, integrity and functioning of ecosystems impacted by climate change, safeguarding of ecosystems that contribute to human health, livelihoods and well-being, and ecosystem-based management of fishery resources. While some of those could be regarded as means towards an end of conserving ecological diversity, others cannot.

Although this application of the established term "diversity" to an ever-widening field has not aided comprehension, the broadened focus does have conservation merit. It should always have been clear that ecological diversity is not merely a simplistic summary of a biotic community but a simplistic summary of just one facet of a community: its Species Abundance Distribution, meaning the relative abundances of the species that it contains, and perhaps equivalent distributions of units larger or smaller than species. Whether seen as biological systems with intrinsic value or as sources of ecosystem goods and services, ecosystems are vastly more than the SADs of the communities they contain, while for upwards of 30 years it has been clear that indices of ecological diversity, *sensu stricto* or *sensu lato*, have little to tell of those aspects of ecosystems which are important, intrinsically or economically. Those metrics fail to capture information on gross productivity, on the conservation status of species at risk or on the extent of anthropogenic degradation. Indeed, ecologists have known (or should have known) that the variety of life, however it is expressed and quantified, is a necessary but not a sufficient focus for the goals of conservation efforts. A much wider perception of the nature and state of ecosystems is needed for their conservation. To a degree, that required breadth has been achieved, without disrupting the (erroneous) public supposition that "diversity" equates to "goodness", through the expedient of broadening scientific notions of the meaning of "diversity" in parallel with the public ones – and in despite of the CBD definition. The resulting confusion has, however, led to research resources being expended on studies of species richness, when the questions to be addressed have concerned quite other facets of ecosystem status.

---

[7] To date, the two "headline indicators" that can be said to directly relate to non-human diversity, have only been implemented using indices based on IUCN Red List rankings of species of mammals, birds or amphibians and similar rankings of extinction risks of mammalian and avian domesticated breeds (BIP 2010).

These shifts in the meaning of words accompanied another important change: Ecological diversity was originally conceived as a description of what is – a state of nature. When that diversity was hypothesized to be a driver of productivity and stability, it became overlain with human values but diversity remained a descriptor of what is or might be. Once the concept passed into the hands of activists and opinion leaders, however, positive trends in biological diversity (or a reduction in negative trends) became transformed into a policy goal of what should be. In Mayer's (2006) terms, diversity changed from being "value-free" into something "valuable". While that fundamental modification should not affect the quantitative values of diversity indices, it does alter their application and perhaps the rigor and precision with which they should be measured.

More generally, ecological diversity was explicitly developed as a step towards simplifying the complexities of ecosystems, part of an attempt to reduce them to a handful of quantifiable metrics. In contrast, biodiversity, as it seems to be widely understood, embraces and even celebrates the complexity. It includes but is not limited to the quantifiable. Thus, Mayer (2006) saw biodiversity as a general concept which can only be defined with reference to specific contexts and "which cannot be caught in numbers", such that only facets of it can be quantified (cf. Sarkar 2010).

The European Union has confronted head on this discrepancy between the narrowness of ecological diversity and the broad array of facets of ecosystems that are important ecologically, economically and for conservation. The EU was led to a new and all-embracing term, "Good Environmental Status", only the first of eleven descriptors of which concerns biodiversity, while only three of fourteen "indicator types" for that one descriptor relate to biotic variability and hence to the CBD definition of "biodiversity" (ICES 2011c; 2012b). Ecosystems are so different, one from another, that any attempt at a general definition of "GES" is certain to be non-quantitative and indeed tautological, while the term is specific to the EU initiative and it would only cause confusion to adopt it outside of that context[8]. None the less, denoting the full breadth of issues that some have sought to place under "biodiversity" as "environmental status" would add clarity and avoid much confusion. Thus, for the purposes of this report we adopt the latter term as including all facets of the condition of ecosystems, reserving "biological diversity" and "biodiversity" for the subset of those facets which fall within the wording of the CBD definition. Abundances, ranges, conservation risks, productivities, mortality rates, stabilities, carbon sequestration and other ecosystem services can all be important aspects of environmental status but none directly address "variability among living organisms" and hence they are not seen here as facets of biodiversity. The same can certainly be said of anthropogenic drivers, including pollution, fishing effort, the anthropogenic component of climate change or various conservation initiatives. Ecological diversity (including richness and evenness), phylogenetic diversity, functional trait diversity, habitat diversity, intra-specific genetic diversity and distributions of individual body sizes all concern variability within biological systems and hence are here classed as facets of both environmental status and biodiversity. We would here consider variability in biological and ecological processes as facets of biodiversity but no metrics of "process diversity" are known to us and hence we do not address that topic.

---

[8] That context has not been free of problems and our recognition of the merits in the idea of "environmental status" should not be taken as a recommendation of the EU's implementation of it.

Beyond the limits of our "environmental status", some writers have made yet another leap by converting "biodiversity" from the variability among organisms into the products of that variability, which is to say the organisms themselves – ultimately equating "biodiversity" with totality of all living forms in the biosphere (Swingland 2001). That seems an entirely unwarranted extrapolation but it has been echoed by such an authoritative body as the United Nations Environment Program, which has declared:

> The term 'biodiversity' is indeed commonly used to describe the number, variety and variability of living organisms. This very broad usage, embracing many different parameters, is essentially a synonym of 'Life on Earth'[9].

While that step from variability in biotic systems to living organisms seems to largely belong in the lay community, albeit drifting into the policy sphere, it is not unknown in the scientific literature. Beaumont *et al.* (2007), for example, sought to identify what they termed the "goods and services provided by marine biodiversity" but then addressed the benefits derived from living systems, not those from the variability in those systems – a very different topic. Likewise, in its report on measuring biodiversity, the Royal Society (2003) declared that it had adopted the CBD definition of the term but then went ahead to discuss the "significance of biodiversity" in terms that could only relate to the significance of living things, without reference to the variability among those things.

For the purposes of this report, we reject all such conflation of variability with the products of that variability. We see nothing but confusion arising from equating "biodiversity" with Life on Earth.

2.2 USES AND INTERPRETATION OF BIODIVERSITY MEASUREMENTS

Measurements of ecological diversity may have applications in ecological research. That was their original intended purpose and some ecologists find that they still have utility in attempts to understand how ecosystems function. In that setting, indices of biodiversity can serve to summarize data on some facet of the structure of a community. Alternatively, some authors have sought to construct ecological theories and used those to predict the SADs that should result, sometimes testing those outcomes against field measurements of diversity. That approach as been used since the 1960s at least (cf. Magurran 2004) but notably since Hubbell (2001) advanced his Unified Neutral Theory of Biodiversity (which concerns what is here termed ecological diversity *sensu lato*: cf. McGill 2010; Rosindell *et al.* 2011). Ecological research is not, however, the focus of the present report, which addresses biodiversity measurement for management purposes, and the complexities of research can be left aside – with a warning that anyone contemplating such work needs to look beyond the recommendations that we offer.

---

[9] http://www.unep-wcmc.org/what-is-biodiversity_50.html (accessed 20 October, 2013)

In the context of work in support of management, there are two broad areas in which measurements of biodiversity might be used. The primary one involves tracking temporal change within a community, including in relation to the CBD's, or other, management targets, through what may be called "trend monitoring". The second area of use lies in mapping biodiversity as a step towards mapping ecosystems, *en route* to decisions concerning spatial planning, such as the siting and design of MPAs. Usually such maps will be drawn relative to physical space (e.g. with axes of latitude and longitude) but biodiversity could potentially be mapped in some other "space". Conceptually, one could examine temporal change in mapped diversity but that would become a three (if not four, should depth be considered) dimensional task and hence is unlikely to be undertaken routinely. Some would see a third application of biodiversity measures in indicators of stressors or anthropogenic impacts on biodiversity (e.g. Heink & Kowarik 2010). However, diversity commonly has a non-monotonic relationship to anthropogenic stress (see Section 2.3 below) and so tracking the latter is better approached by combining measurements of anthropogenic activities with descriptors of spatio-temporal variations in some facet or facets of biodiversity – which is to say that the required measurements of diversity fall into the categories of trend monitoring and mapping. Hence, this report primarily addresses biodiversity measurement as a foundation for those two tasks.

When mapping biodiversity as a support for conservation planning, the ideal product might be a map of what diversity would be at some future time (how far into the future depending on the planning horizon) in the absence of anthropogenic effects or with those effects concentrated into areas not under consideration for closure. All that can be mapped in practice, however, is the realized diversity (including any anthropogenic effects) at the time that the survey data were gathered. It perhaps bears repeating that if a map of diversity free of human impacts could be drawn, it would not display uniformity: natural biodiversity is spatially variable. It follows that a real map cannot be simply interpreted in terms of anthropogenic impact or of "wildness". Areas of high or low biodiversity at present result from a combination of natural spatial variations, spatial variations in anthropogenic pressures and those in the responses of local biodiversity to the pressures. The natural and anthropogenic causes cannot be separated with the information generated by a single survey, though the separation might be approached through more complex experimental designs.

In trend monitoring, it is the realized diversity at the time of the survey that is of interest, though the focus is on temporal change from one survey to the next.

Ecologists well understand that stasis is an unusual condition for any ecosystem but early simplifying assumptions, involving equilibrium conditions, have affected the terminology used and biodiversity change is often considered in terms of a response to "disturbance" (e.g. Dornelas *et al.* 2011). It is explicitly understood that that can be either natural or anthropogenic disturbance, with changes in realized diversity reflecting the results of both working in combination. Hence, detected temporal trends in biodiversity cannot simply be assumed to represent increases or decreases in either human impacts or their converse: "wildness" (Green & Chapman 2011). As with consideration of spatial patterns, examining causality requires appropriate experimental design. That must include appropriate temporal scales. Long-term trends may reveal much about anthropogenic impacts but the latter may themselves be subject to major short-term fluctuations in response to management actions, while decisions on future management should be founded on careful, balanced comparisons between near-current conditions and those before substantial anthropogenic degradation (cf. Nicholson & Jennings 2004).

2.3 THE MEANING AND RELEVANCE OF ECOLOGICAL DIVERSITY

The environmental status of an ecosystem is important, though the reasons are tautological: any aspect of an ecosystem that is deemed important can be considered to be a facet of its environmental status, as that term is used here. The ultimate foundation of biodiversity (and still, for many, its core) is, however, ecological diversity. The latter began as a way of briefly summarizing a Species Abundance Distribution and that it successfully does – at the cost of discarding much of the information in the SAD, along with the (typically expensively obtained) names of the species and any additional data that may have been gathered (e.g. individual sizes of the organisms or indications of their diets). That discarding can have serious consequences. As Leinster and Cobbold (2012) put it, to use indices of ecological diversity is "to ignore the plain fact that some species are more similar than others. Deliberately ignoring biological reality is unlikely to lead to a helpful assessment of diversity". Whether ecological diversity has any deeper or broader meaning or relevance than merely being a summary of a SAD is much less sure. Pielou (1975), for all of her emphasis on the topic, certainly did not over-state the depth of diversity measurements, declaring: "a community's diversity index is merely a single descriptive statistic, only one of the many needed to summarize its characteristics and, by itself, not very informative".

Over the decades, many ecologists have used diversity metrics in research work, justifying such applications as methodological choices must always be justified. Those uses, however, lie outside our present concern over trend monitoring and mapping in support of management. It is meaning and relevance in the latter roles that are of immediate interest and either is hard to find.

As has been said above, ecological diversity *sensu stricto* is blind to absolute abundances, while species richness ignores abundances altogether. Index values can remain stable despite severe depletion, provided that relative abundances remain unchanged. Ecological diversity does not respond to species substitutions, if the distribution of relative abundances within the community (not necessarily the relative abundances of particular species) does not change. While exact constancy of relative abundances is not to be expected during major species substitutions, nor indeed across periods when there is only minor change in species composition, evenness is as likely to increase as it is to decrease when ecosystems undergo upheaval, making ecological diversity a very poor guide to change, save when a community is so severely degraded that richness is substantially reduced. No amount of thought devoted to improved indices, nor conceptual advances, can overcome these limitations. They were and are inherent to the choice, made between the 1940s and 1960s, to focus on Species Abundance Distributions and to reduce each SAD to a single scalar number or at most three of them. That choice remains fundamental to the basic conception of ecological diversity.

The common belief that ecological diversity is inversely related to anthropogenic impact, and hence directly to "wildness", can be swiftly dismissed. Since the 1970s, ecologists have been familiar with the Intermediate Disturbance Hypothesis, which suggests that moderate levels of disturbance (across small spatial scales) increase species richness by removing climax species and opening opportunities for others (Rosenzweig 1995). While that can apply to the effects of natural stressors, it is equally applicable to many kinds of anthropogenic disturbance (Dornelas *et al.* 2011). The effect is easily visualized in forestry contexts: driving a logging road through a stand of trees will increase angiosperm richness, on the scale of the stand, by allowing colonization of the road margins by low plants that would have been shaded out by the trees. In an aquatic context, one of the most obvious anthropogenic losses of ecological diversity results from severe eutrophication of small lakes and even some coastal marine waters. In extreme cases, all metazoan life can be eliminated. Yet many Canadian glacial lakes are naturally oligotrophic and in such waterbodies moderate eutrophication increases diversity (cf. Magurran 1988). Hence, ecological diversity often bears a non-monotonic relationship to the degree of anthropogenic disturbance. Without extensive research to discover the amount of some human activity that corresponds to maximal diversity in a particular system, ecological diversity is therefore useless as a measure of "wildness".

That is not to say that measures of ecological diversity can never be used to rank communities by their degrees of anthropogenic degradation. In particular situations, that can certainly be done. It requires, however, that all of the compared communities be subject to the same principal stressor and that all of them are stressed beyond the intermediate point that produces maximal diversity. Thus, meiofaunal diversity in urbanized harbours can indicate the relative severity of hydrocarbon pollution (e.g. Moreno *et al.* 2008) and similar measures can reveal degradation close around offshore oil platforms (e.g. Olsgard & Gray 1995). However, the non-monotonic relationships between anthropogenic pressures and the resulting effects on ecological diversity prevent the latter from being used as a general indicator of the former.

It should not need to be said that no metric of ecological diversity can be used as a simple, quantitative measurement (as in a "measuring stick") of anthropogenic impacts, to be applied across multiple ecosystems. No metric can fill that role because ecosystems, and the taxocenes within any one ecosystem, differ very markedly in their natural diversities and in their responses to various anthropogenic pressures. Nor are those differences static in time: following disturbance events, be they anthropogenic or natural, the process of ecological succession usually produces an initial, marked increase in species richness until an asymptote is reached, after which some ecosystems show a moderate decline (Rosenzweig 1995). There should be no need to say that no such metric can serve as a "measuring stick" and yet the European Commission has produced a Guidance Document (EC 2003) in which a Norwegian "classification tool" for "soft bottom macrofauna" is presented without critical comment. That "tool" uses values of Shannon's $H' > 3$ to indicate "good" or "very good" ecological status, while values < 2 indicate "bad" or "very bad"[10] . Marques *et al.* (2009) offered an earlier version of the same "tool". Any such interpretation of H' is highly suspect when applied to a particular dataset and ludicrous if extended to other data. Yet Zettler *et al.* (2007) took the EC (2003) presentation as a proposition of "an absolute scale [...] of EcoQ".

The other ecological meanings of ecological diversity, first suggested a few decades ago, are that it is linked to both enhanced productivity and stability of communities or ecosystems – though the linkage need not be directly causal. That is: higher ecological diversity may be a consequence of some other community characteristic which leads to enhancement of productivity or stability (McCann 2000). Even if the linkage is causal, the direction of causality cannot be assumed to be one of diversity influencing productivity: much early research went into the effects of primary productivity on species richness in natural systems, finding that the direction of change in richness, in small scale experiments, depended on the factors driving increased productivity, while regional richness usually bears a non-monotonic dependence on primary productivity (Rosenzweig 1995).

Empirical support for the idea that enhanced ecological diversity drives greater productivity and stability was very thin until the last twenty years. Following the advent of the CBD, there have been several hundred experimental studies which are beginning to provide an understandable picture of the effects, though most of the work has been on plant production in agricultural settings. Besides that limitation, even hundreds of experiments are entirely insufficient to cover all of the combinations of communities, levels of diversity, forms of production and stability, as well as scales of time and space that may be of interest somewhere in the biosphere. A major difficulty in such research is that, when working with highly complex ecological systems, only a controlled, replicated, manipulative experiment can demonstrate the causation of observed trends in diversity, production and stability. Yet such experiments are highly challenging when investigating large, mobile animals and, even with plants, are impractical to the point of being impossible at the scales of whole ecosystems or over the long time periods across which diversity loss is expected to have its deleterious consequences. Hence, most experiments to date have concerned either small or immobile organisms and have been limited in space and time to scales which may have limited relevance (while, as ever in ecology, fundamentally different processes can dominate at different scales: Rosenzweig 1995). Only a subset of the mechanisms through which diversity may act have been investigated (Stachowicz *et al.* 2008). Meanwhile, the reductionist approach implicit in experimentation tends to miss the point of relationships within complex ecosystems and perhaps leads to underestimates of the effects of ecological diversity (Hillebrand and Mattiessen 2009). McCann (2000) likewise doubted that experiments focused on single trophic levels could ever be sufficient since, in his view, it is the diversity of linkages between levels that is key.

---

[10] The available on-line version of EC (2003) lacks the quantitative values of H' that define the classes of ecological status, though it presents the same idea with colour coding. The missing quantitative values can be found in the draft report cited by Zettler *et al.* (2007).

Within those limitations, a number of studies have shown that primary production by angiosperms increases with species richness of the plants, though it reaches an asymptote within about a dozen species and sometimes with just one or two. The richness of trait groups (each containing species that share common responses to the environment or that exert similar influences on ecosystem properties) proves to be more important to production than does species richness – meaning that the diversity of "redundant" species, those which might replace one another, has little effect on production. Even within the narrow topic of the distribution of abundances across species, community composition (meaning the particular species that are present with varied relative abundances – to which ecological diversity is blind) has a greater influence on production than do SADs and the ecological diversity calculated from them (Hooper *et al.* 2005). Duffy (2009) has argued, however, that those conclusions arise from the limitations of the experiments. Maximizing the full suite of goods and services from a spatially-extensive ecosystem, and maintaining that maximum over time, will need much higher species richness than simply maximizing one output (e.g. primary production) in a short-term experiment on one small plot, since it can be expected that different species will support different ecosystem services or maximize the same service at different times.

Recent reviews have disagreed over the hypothesized increase in production at higher trophic levels that is supposed to be driven by species diversity. Hooper *et al.* (2005) concluded that it has not emerged in experiments, the functional characteristics of the species present having a greater effect than ecological diversity. In contrast, a meta-analysis by Balvanera *et al.* (2006) found evidence of an increase in animal production with greater diversity. Part of the disagreement may lie in the choice of which prior studies to consider in reviews. Many experiments that involved manipulation of plant diversity have looked for enhancement of herbivore production and found little, since the positive effects of greater ecological diversity at one trophic level appear to be muted at others. The few studies that have investigated the effects of herbivore diversity have suggested that it does enhance herbivore production (Balvanera *et al.* 2006), even if plant diversity does not.

Meanwhile, a number of experiments have indicated that higher within-site ecological diversity acts to increase a community's resistance to invasives, though sites with higher diversity are more vulnerable (Hooper *et al.* 2005). While the full causes of that apparent contradiction are doubtless complex, it may be more understandable if seen in terms of sites with a greater variety of niches offering more opportunities for invasives to prosper, whereas those sites with all niches already occupied by multiple effective competitors have the opportunities closed off. The relevant experiments reviewed by Hooper *et al.* (2005) were all conducted in terrestrial ecosystems, however, and there is no certainty that aquatic ones are regulated in the same way – an offshore benthic community limited by the supply of planktonic larvae might respond quite differently to a space-limited inter-tidal community, for example, let alone to how terrestrial ones behave.

The most prominent effect of ecological diversity on ecosystem properties is the tendency for higher diversity to lead to greater stability (cf. McCann 2000; Stachowicz *et al.* 2007), though that has been a very contentious conclusion. Much of the difficulty has arisen from the multiplicity of meanings of "diversity" and even more those of "stability" (Ives & Carpenter 2007). Certain kinds of stability in the emergent properties of ecosystems do appear to be enhanced by higher diversity but not all forms of elevated diversity have that effect. McCann (2000) suggested that it is functional diversity, rather than species diversity, that enhances ecosystem stability, whereas Hooper *et al.* (2005) stressed the enhancement by "redundant" species, meaning ones that are functionally interchangeable (in contrast to the positive effect on production, where the number of functional groups is important but the number of species within each group is not). Given redundancy, as some species are depressed (whether by rising temperatures, anthropogenic discharges or some other factor), others can expand and maintain the overall production, or other property, of the ecosystem. The apparent conflict between McCann's (2000) and Hooper *et al.*'s (2005) positions may be merely one of human perceptions: A primary producer and a top predator are functionally different but so different that they cannot replace one another and hence their mutual presence cannot enhance stability. Indeed, ecosystems with a greater variety of species filling unique roles have more points of potential instability. In contrast, two congeners that are barely distinguishable by taxonomists may have such similar responses to perturbation that they cannot substitute for one another. Thus, ecosystem stability may be enhanced by the diversity of groups of species that are similar enough to be substitutes but different enough that one can thrive when another is depressed. Perhaps fortunately, species are rarely entirely redundant, such that substitution can occur. Redundancy is, however, usually less developed at higher trophic levels, such that depression of one predator species can lead to major ecosystem changes (Hooper *et al.* 2005).

Combining ecological theory with empirical observations, McCann (2000) advanced the idea that the stability-enhancing driver is the presence of weak links among species, which serve to damp the destabilizing effects of strong links. A species with only one prey type and only one predator, for example, would promote a destabilizing trophic cascade if some external factor depressed the predator, whereas species with broad prey preferences that are preyed upon in their turn by multiple predators would tend to add stability. Thus, it is the typical structure of food webs, with many weak links and few strong ones, that stabilizes ecosystems (McCann 2000). If so, higher diversity is not the cause of the stability enhancement but rather a consequence of there being more species sharing the total linkage between each pair of adjacent trophic levels. Any anthropogenic decline in diversity would, nevertheless, reduce stability as those linkages came to be primarily shared between fewer abundant species (McCann 2000).

Whatever the mechanisms, the enhanced stability does not emerge in community responses to all stressors. It is seen in experiments concerning nutrient perturbations but not those which consider climatic warming, for example (Balvanera *et al.* 2006). Most seriously, the "stability" is in the emergent properties of the ecosystem, such as overall primary production, not in its community composition (McCann 2000). Indeed, stability of production comes through instability of species abundances, with one replacing another as conditions change (Hooper *et al.* 2005) and, in their meta-analysis, Balvanera *et al.* (2006) found evidence for an increase in compositional instability with higher diversity.

Such stability of emergent properties has long been familiar to fisheries ecologists. It was perhaps first observed in the Laurentian Great Lakes, where fishery production remained steady despite major changes in fish community composition (Smith 1968; Kelso *et al.* 1996), and similar observations have since been reported from other aquatic systems. While trends in real ecosystems cannot be explained by single causative factors, the effects of diversity on the stability of community properties can be illustrated by comparing the groundfish fisheries off northeast Newfoundland with those on Georges Bank. When the cod (*Gadus morhua*) in the former area declined under a combination of high fishing pressure and an unfavourable environment, around 1990, groundfish landings collapsed. The bottom fisheries of Georges Bank, in contrast, initially targeted halibut (*Hippoglossus hippoglossus*), then cod and later haddock (*Melanogrammus aeglefinus*). As those species were fished down, flatfish (especially yellowtail flounder: *Limanda ferruginea*) became more important. They were followed by a brief period of abundant dogfish (*Squalus acanthius*) and then one when a mixture of skates was prominent (e.g. Fogarty & Murawski 1998). The emergent property in that comparison, overall groundfish harvest, remained remarkably steady throughout on diverse Georges Bank but collapsed in the low-diversity northeast Newfoundland area. The contrast argues for a strong, positive effect of ecological diversity on the stability of ecosystem goods and services. The nature of that stability is not, however, one sought by current policy: very considerable management efforts, backed by strident demands in the public arena, have sought and are continuing to seek stability in production from individual fishery-resource populations and not an exchange of prime cod and haddock for dogfish and skates. Outside of fisheries concerns, conservation programs likewise usually seek stability in species compositions, in addition to stable emergent properties of ecosystems, rather than purchasing the latter through sacrifice of the former – doubly so when the species in question is deemed to be “at risk”. That desired species stability does not appear to be enhanced by greater ecological diversity.

Most of the experiments which have shown this enhanced stability have used species richness, rather than ecological diversity *sensu stricto*, as their independent variable (Hooper *et al.* 2005). The implications for stability of changes in evenness are much less clear, while Ives and Carpenter (2007) claimed, without citing evidence, that it is often the presence of one or a few particular species that confers stability in the face of a specific stressor, rather than overall diversity.

Such broad generalizations are apt to be misleading, however, and various types of ecosystems may respond quite differently. Balvanera *et al.* (2006) found that the positive effects of higher diversity on ecosystem properties are small in the most commonly studied types of ecosystems, including marine, freshwater, grassland and forest systems, but considerably larger in saltmarsh, bacterial and soil ecosystems. They also found that the effects are less in field experiments than in those more closely controlled, suggesting that environmental heterogeneities may mask much of the benefit of higher diversity. Since real ecosystems are invariably heterogeneous, the masking may be important in conservation applications, even if it is a distraction to academic study.

For marine systems in particular, Stachowicz *et al.* (2007) recently reviewed the empirical evidence for effects of ecological diversity. They drew on a total of 123 experiments, many undertaken in communities containing macrophytes and hence concerning the inter-tidal or shallow-sublittoral benthos. Perhaps the major conclusion from their work is that “identity effects”, the effects of particular species on community properties, are both greater and more nearly universal than are the effects of richness – much as suggested by Hooper *et al.* (2005) and Ives and Carpenter (2007) for other ecosystems. In the sea, diverse communities do tend to perform better than the average of monocultures of the same species, though not as well as a monoculture of the single best species. Hence, the advantages of diversity emerge when multiple aspects of community response are considered together, different particular species being especially important to different properties (Stachowicz *et al.* 2007). It is commonly seen that the most productive benthic species are not the ones that dominate in polyculture. Greater diversity can dampen the top-down effects on ecosystems of species at higher trophic levels but other across-trophic-level effects are less clear, with different experiments showing contradictory results. Greater predator diversity can either increase or reduce plant biomasses, for example, depending on whether or not the predators eat one another as well as the herbivores (Stachowicz *et al.* 2007).

In summary, ecological diversity is not the powerful indicator of environmental status that it is often represented to be. While it can be negatively affected by human impacts, that relationship is too far from being universal, or universally monotonic, for diversity to be used as a measure of "wildness" or "naturalness". Greater ecological diversity can enhance production but usually not (in the short term and in marine systems) by as much as correct selection of the optimal species for monoculture would provide, while trait diversity appears to have a greater effect than species diversity. The benefits of the latter are seen primarily at the trophic level which is more diverse, being swiftly dissipated up or down the trophic pyramid. Ecological diversity does appear to be linked to community stability across a variety of ecosystems but it is not clear that that is a form of stability which is of much interest in conservation management. Nor is the effect driven by species diversity *per se* but rather by the diversity of redundant species. Throughout, community composition, meaning the particular species represented by the relative abundances in a SAD, proves to be important but that is an aspect of field data that is entirely discarded during the evaluation of ecological diversity. Some of this generally-negative conclusion has arisen from the limitations of experimental investigation. As Ives and Carpenter (2007) have argued, even if the continued maintenance of ecosystem goods and services, in the face of ever-shifting pressures, depended on single key species and single genes within those species, rather than on diversity, the species and even more the genes that confer stability when an ecosystem is stressed will almost certainly depend on the stressor, or combination of stressors, in question. Without an impossible program of experimental investigation of every conceivable stressor and every individual ecosystem on which it might bear, the key species and genes cannot be known and hence the only rational management action is broad-scale conservation of biodiversity.

What does emerge from the empirical evidence is that, while ecological diversity is (comparatively) easy to measure and hence to monitor, it is a poor surrogate for those ecosystem properties which matter, such as productivity and stability. Forty years ago, R.W. Poole suggested that measures of ecological diversity are "answers to which questions have not yet been found" (Ricotta 2005). That went too far but, while richness, evenness and ecological diversity *sensu stricto* may continue to be useful measures in some kinds of ecological research, they are rather meaningless for management purposes and those ecosystem properties that are important to environmental status should be monitored directly. It is rarely sufficient, nor even very useful, to determine a value for ecological diversity as an indirect indicator of that status.

2.4 BIODIVERSITY CHANGE IN THE SEA

While that conclusion can stand for all ecosystems, matters are more extreme in the sea. The study of anthropogenic declines in ecological diversity has primarily focused on terrestrial settings, where humans transform ecosystems through urbanization, deforestation and other fundamental changes. Conservation biologists are thus much concerned with loss of species (both global extinctions and local or regional extirpations) and with their gain through invasions or introductions (e.g. Hooper *et al.* 2005). Gotelli and Colwell (2011) thus saw the number of species remaining in a community as "the ultimate 'scorecard'" of conservation. Species loss and gain are also very serious issues in freshwaters, particularly the smaller ones though even the ecosystems of the Laurentian Great Lakes have been profoundly affected by invaders. Similar changes in species richness can be of concern in semi-enclosed marine basins, such as harbours and lagoons, and to a lesser degree in near-shore, coastal waters. However, despite promotion of the extinctions / introduction metaphor as being applicable to all marine ecosystems (e.g. Byrnes *et al.* 2007), the open sea, away from the immediate influence of the coastline, is quite different.

2.4.1 Extinctions, Invasions and Extirpations

Contrary to claims widely disseminated through the mass media, there has not been even one recorded global extinction of any species that is capable of completing its lifecycle in the open sea during historical times (and so potentially with an anthropogenic cause), if by "open sea" is meant waters more than a few miles from the nearest land. Dulvy *et al.* (2003) found 21 examples of what they considered to be global marine extinctions in recent centuries (a count later revised to 20: Dulvy *et al.* 2009), plus a further 112 regional extirpations. Dulvy *et al.*'s (2003) list has, however, been dissected by Monte-Luna *et al.* (2007), who concluded that only sixteen global extinctions can be confirmed, along with 50 extirpations. Others in the original compilation were mostly cases of extreme depletion that have yet to reach extirpation, while some may not even be examples of marked depletion at all (Monte-Luna *et al.* 2007). Of the 21 global extinctions claimed by Dulvy *et al.* (2003), seven were terrestrial-breeding mammals or birds (expanded to eleven by Dulvy *et al.* 2009), one was an anadromous fish, two were macroalgae with reported ranges confined to a single urbanized harbour (reduced to one by Dulvy *et al.* 2009), eight were intertidal or shallow sub-littoral benthic invertebrates (the extinction of one of which, the Californian white abalone, *Haliotus sorenseni*, has not yet occurred, while the status of others has been disputed: Monte-Luna *et al.* 2007; Dulvy *et al.* 2009), while two were reef fish of near-shore waters (though the status of one is disputed: Monte-Luna *et al.* 2007). The only other known marine global extinction was Steller's sea cow (*Hydrodamalis gigas*), a near-shore species so vulnerable to hunting that it was eradicated from most of its native range (California to Japan) in the prehistoric era. The sea cow avoided being counted amongst the Pleistocene megafaunal extinctions only because it survived into the mid-18th Century around the uninhabited (by humans) Commander Islands and some thinly populated parts of the Aleutians (Anderson 1995; Domning *et al.* 2007; Turvey 2009).

There may have been some as-yet unrecognized extinctions in the open sea since Dulvy *et al.* (2003) found an average lag of 53 years between the last sighting of a species and recognition of its extinction or extirpation. While that result was partly an artifact of the timing of modern conservation concern, some losses of recent decades likely await recording. It is also probable that some marine species unknown to science have been lost, with their passing necessarily unnoticed (though the loss of unrecognized species cannot alter recorded diversity). The complete lack of known global extinctions away from the coast suggests, however, that the unknown ones have been very few – perhaps confined to weaker members of pairs of "cryptic species" (a stronger congener having replaced a weaker without taxonomists recognizing that there was ever more than one species) and to cases of extreme endemism, itself rare in the sea. It has been postulated, for example, that some endemic benthos on seamounts may have been lost to destruction of their habitats by bottom trawling (e.g. Samadi *et al.* 2007), though the degree of seamount endemism has since been disputed. In short, while the anthropogenic global extinction of any species is unacceptable, the number of such losses in the open sea cannot be more than a negligible contributor to biodiversity change.

Meanwhile, the pattern of species invasions in the sea differs from that on land because of the immensely greater dispersal abilities typical of marine species – and does so despite some terrestrial species sharing those abilities while some marine ones lack them. As a first approximation, it can be assumed that any species capable of surviving in a particular area of open sea found its way there naturally in pre-modern times. Exceptions are known but they are few. Most obviously, the distribution of the continents severely limits natural exchange between the Pacific / Indian basin and the Atlantic / Arctic / Mediterranean one, save at high latitudes. Where those barriers have been breached or surmounted, invasions have occurred – "Lessepsian" invasions from the Red Sea into the eastern basin of the Mediterranean *via* the Suez Canal being notable (e.g. Por 1971; Golani 1998). In a very different example, live release into United States Atlantic waters of Indo-Pacific lionfish (previously purchased for home aquaria) seems to have led to establishment of a self-sustaining population (e.g. Whitfield *et al.* 2002). Ballast water transport of propagules, a serious problem for coastal and inland aquatic ecosystems, appears to resulted in a few invasions of offshore areas, such as the tunicate *Didemnum* sp. on Georges Bank (Lengyel *et al.* 2009).

What have been seen amongst open-sea species are regional and local losses. Unfortunately, their nature and extent are obscured by terminological confusion. "Extirpation" (sometimes called "local extinction": e.g. Dulvy *et al.* 2003) should mean the elimination of all individuals of a species from some definable area but that term is sometimes used for the loss of a semi-discrete population (e.g. Kenchington 2003) – an elimination that is at once local, since any one population must have a restricted range, and yet global, since the loss of a population from its own locality is also its loss from the biosphere. For aquatic species in isolated lakes, as for terrestrial species on isolated islands, those two forms of local loss are often synonymous. For typical marine species, in contrast, they can be entirely unrelated. In many fish, for example, the range of local, coastal "stocks" is entirely occupied at some point in the year (outside of brief spawning seasons) by larger, migratory populations. Hence, complete elimination of local populations could occur with no reduction in species range whatever. Conversely, migrant species that move inshore to feed can opt not to enter particular areas if they become unattractive (e.g. because of silt loads, hypoxia, or an absence of prey). In that event, there can be a detectable range contraction without loss of any local population. For clarity in this report, we use "extinction" or "global extinction" to mean the entire loss of a species or other described taxon, "extirpation" to mean the loss or withdrawal of a species from a defined area, and "local-" or "population extinction" when it is necessary to refer to the loss of some semi-discrete population that has not been formally described as an infra-species or sub-species. A single event may constitute one, two or all three of those simultaneously.

In contrast to global extinctions, there have been multiple local extinctions and extirpations in open-sea environments in recent centuries. Dulvy *et al.* (2003) listed only the loss of one entire baleen whale population, the East Greenland bowhead (*Balaena mysticetus*), and the extirpation of various elasmobranchs from particular portions of European waters (the evidence for most of the latter being highly questionable: Monte-Luna *et al.* 2007). However, there have been many other examples, beginning with the Atlantic grey whale (*Eschrichtius robustus*) in the 17th or early 18th Centuries if not before. An example of note in Atlantic Canada is the (as yet unexplained) withdrawal of haddock from the Gulf of St. Lawrence through the course of the 20th Century, which may now have led to their complete absence from that sea area (Kenchington 1996). Interpretation of regional extirpations is, however, complicated by the variability in marine species' ranges that is linked to ocean climates. The multi-decadal "Russell Cycle" in the zooplankton of the western English Channel may have been the first example described (cf. Robinson & Hunt 1986). Of far greater magnitude and significance are the changes in the Pacific basin driven by *El Niño* events, which affect the biota off the west coast of Canada (e.g. Pearcy & Schoener 1987; Ware 1995) as well as elsewhere. Other known, if lesser, temporal changes include the various eruptions of warmer-water species into the North Sea (e.g. Reid *et al.* 2003) and the varied penetration of Labrador Sea Water at upper continental-slope depths, which can be limited to the western side of Grand Bank but may extend as far as Maryland (MERCINA Working Group 2001; Greene & Pershing 2003), with apparent effects on biodiversity (Marsh *et al.* 1999). On a still smaller scale, populations of various species are known to undergo range contractions when depleted (e.g. Shackell & Frank 2003; Cardinale *et al.* 2012) but are expected to expand again if their abundances recover. All of these distributional changes will appear as "invasions" or "extirpations" if considered at spatial and temporal scales smaller than those of the shift in ranges.

A more serious form of biodiversity loss is the relative depletion, and even local extinctions, of minor populations of some resource species. Canada has long managed most of its fisheries in terms of multiple "stock" units within each species, perhaps leading the world in that regard, but for almost as long there has been evidence of population structuring at far finer scales than those of the established "stocks" (cf. Kenchington 1984; Smedbol & Stephenson 2001). Atlantic cod is particularly noted for such subdivision: Evidence for genetic distinctions between its local groups is fragmentary (but see Ruzzante *et al.* 1996, 2000) but the former existence of semi-discrete inshore populations is well attested (e.g. Kenchington & Halliday 1994; Campana *et al.* 1995; Smedbol & Stephenson 2001; Ames 2004). It has recently been suggested that haddock once had a similar structure (Fowler 2011). While firm evidence is generally not available, a suspicion lingers that fisheries have tended to deplete less-productive or more-available local populations, relative to those larger units which tend to dominate the biomass estimates of stock assessments, resulting in a decline in intra-specific biodiversity. Adequate mapping of population structure has never been attempted and hence many semi-discrete units may have been lost without their existence being recognized. In the short term, that has perhaps had more effect on human interests (notably those of small-boat fishermen, dependent on local fish runs) than on ecosystems. Over the long haul, however, any narrowing of the variability within a resource species risks compromising its ability to adapt to changing environmental conditions.

Recorded species richness, rather than the biological reality, in the open sea can also be affected by the widespread presence of expatriates (individuals living in places where their species cannot reproduce). Those can be abundant enough to support substantial fisheries, as the West Greenland cod resource sometimes has (Buch *et al.* 2004), or they may be scattered juveniles drifting through and headed for an early death. The latter are frequent on the Scotian Shelf, for example (Markle *et al.* 1980). It follows that regional lists can contain many names of species that are only recorded when a scientist happens to deploy appropriate sampling gear at the right time and place to capture one of the passing specimens. The absence of subsequent records, which could extend over decades, may appear as "extirpation" but has more to do with the availability of research funding.

In summary, the characterization of biodiversity change as a combination of extinctions, invasions and introductions, along with the idea of species richness as a "conservation scorecard" (cf. Hooper *et al.* 2005; Gotelli & Colwell 2011), may be applicable to both terrestrial ecosystems and aquatic ones from the heads of streams out to the coastline but it misrepresents the nature of change in the open sea. There, global extinctions, invasions and introductions (while individually serious) are negligible contributors to trends in biodiversity. Extirpations can be important but can be hard to distinguish in field data from shifts in ranges driven by assorted causes. Extinctions of local populations may have had very serious consequences but can be entirely undetectable through measurements of species diversity.

2.4.2 Depletions and Disruptions

If the sorts of biodiversity loss seen in terrestrial ecosystems rarely occurs in the open sea, it does not mean that marine ecosystems are in a natural state. Indeed, they have been transformed by the consequences of human activities, particularly fishing – which is much the most spatially-pervasive anthropogenic pressure in the sea, ship noise perhaps excepted. Any surviving pristine areas are hard to find. The massive changes have not, however, been in the form of extinctions and extirpations but rather through depletions of once-abundant species. Optimal sustainable fishing is expected to reduce its resource by about half, while too-frequent overfishing has further depleted many populations. Some non-resource species, particularly those long-lived or with low reproduction rates, are also depressed when taken as bycatch. However, most species that are not targeted by any fishery are expected to respond to the depletion of the resources through ecological shifts – predators of depleted species perhaps declining while their competitors and prey increase. Such changes in relative abundance do not affect richness (save locally, if they drive range contractions or expansions) but they might be reflected as changes in evenness and thence in ecological diversity *sensu stricto*. Hence, it is in evenness, not richness, that the principal signal of anthropogenic impacts on open-sea ecosystems should appear. Unfortunately, the direction of that change is unsure: Most major fisheries target naturally-abundant resources, since few rarities are valuable enough to support economically-viable harvesting. Initial depletion of those resources might be expected to increase evenness, and hence ecological diversity *sensu stricto*. Replacement of the individually-large fish targeted by typical fisheries with an equal biomass (but greater numbers) of a smaller competitor could, however, drive evenness downwards. Thus, the effects of the fishery depletions on ecological diversity may be unclear and difficult to recognize in field data. In practice, while some trends towards increasing evenness under fishing pressure have been seen in the North Sea, other observations have been inconsistent (ICES 2005; Piet & Jennings 2005).

The depletion of fishery resources is not uniform across their size distributions but falls primarily on the largest size-classes. Contrary to frequent claims that the bias results from size-selective targeting (e.g. Shin *et al.* 2005), it is primarily a consequence of the indeterminate growth typical of fishes: fish harvesting inevitably increases mortality rates and, in an established fishery, older age-classes have been exposed to the higher rates for longer, causing them to be depleted further than younger 'classes are. Since older fish average larger, as a result of their indeterminate growth, the consequences of prolonged elevated mortality emerge as a severe reduction in numbers of big fish, even under "optimal" fishing. Adding to the complications, most teleosts change their diets as they grow. Hence, loss of the largest individuals of what were once abundant resources could have severe consequences for the trophic structures of ecosystems. Those consequences, extending over much of the world ocean, likely represent a major loss of global biodiversity but it would have little effect on species diversity, since smaller fish of each resource remain relatively abundant, nor on those metrics of trophodynamic diversity that ignore within-species variation in diets, as many do.

2.4.3 Loss of Genetic Diversity

Some of the above forms of biodiversity loss will necessarily be reflected in a decline in genetic diversity. Global extinctions involve the irrevocable loss of the entire genome of the species in question but that is at least a recognized issue. In contrast, loss of intra-specific genetic diversity is both pervasive and serious in marine as much as in freshwater ecosystems, yet it is too often passed unnoticed. Whereas elevated species diversity has only limited advantages in the short term (see Section 2.3 above), enhanced intra-specific genetic diversity has major advantages in the long term since it is the raw material on which natural selection acts to maintain the fitness of a species in the face of environmental change. Narrowing of that diversity limits the ability of a species to adapt to change, not just over long evolutionary timescales but also much faster, as in the event of a major epizootic or a swift anthropogenic habitat change. From a human perspective, maintaining broad genetic diversity also preserves future options for utilizing genetic resources (Kenchington & Heino 2002).

Few species are truly panmictic. Outside of isolated lakes, however, aquatic species rarely form multiple truly isolated populations. Rather, the normal (genetic) population structure is one of partially-isolated populations, with some degree of exchange of individuals, the amount of genetic difference between pairs of populations typically increasing with the geographic distance between them (Kenchington & Heino 2002). Intra-specific genetic diversity is expressed partly between such populations and partly between individuals within them (as well as in the form of heterozygosity within individuals). The relative proportions which are expressed between and within populations are highly variable from species to species (Kenchington 2003).

The all too common eliminations of local populations may involve irrevocable loss of alleles, if those were unique to the populations lost, or may make some alleles' subsequent loss more likely, if they are rare in other populations (Kenchington 2003). Moreover, some of the between-populations portion of genetic diversity represents adaptation to local conditions, including co-adaptation of different genes – meaning that specialist alleles work together to raise local fitness (Kenchington & Heino 2002; Kenchington 2003). If a population is depleted, relative to adjacent major populations of its species, the rate of immigration will rise relative to population abundance, while mating behaviours that serve to separate the populations at spawning time may break down, both processes serving to increase among-population hybridization. Local adaptation can be lost, decreasing local fitness and hence the species' ability to efficiently utilize resources in every part of its range. Meanwhile, hybrids that chance to inherit alleles for some co-adapted genes from one population and those for other genes from another population may be less fit than pure-bred individuals from either population would be.

Different issues arise when anthropogenic pressures shape selection. Moderate levels of chemical pollutants will encourage selection of alleles that strengthen tolerance of the chemicals in question, for example, though the scope for genetic adaptation will depend on the availability of appropriate alleles within the affected population. In the open sea, however, the primary anthropogenic factor affecting genetic diversity is probably selective fishing. Unfortunately, it tends to change resource populations in directions that are undesirable from the perspective of the human interests – a process termed "Fisheries-Induced Evolution" ("FIE": Laugen *et al.* 2012; van Wijk *et al.* 2013). Almost all fisheries are size selective to some degree, while resource-conservation management tends to increase that selectivity through attempts to protect small fish and hence to prevent growth- and recruitment-overfishing. Intensive fishing for individuals larger than some particular size tends to preferentially remove fish that carry genes for rapid juvenile growth and large sizes at sexual maturity, leaving a disproportionate share of the effective spawners as individuals that show slow growth and early maturity – though the effect on allele frequencies in later generations depends on the heritability of the selected traits, which has been shown to be high in various finfish and shellfish, though the experiments have mostly been done on species considered for aquaculture and in the context of controlled breeding programs. In wild populations, the observed changes in growth and maturity appear to be reversible when fishery selection is eased (Kenchington & Heino 2002; Kenchington 2003).

At least in theory, FIE could operate in other directions, depending on the selection process in a particular fishery and, to some degree, the management measures applied. Reversing such effects, should they occur, could take thousands of years, since natural selective pressures are much weaker than those imposed by intense fishing (Laugen *et al.* 2012).

2.4.4 Summary

Thus, biodiversity change in the open sea has little to do with species extinction or invasion, while observed extirpations are often responses to shifts in ocean climate or trends in abundance. Major anthropogenic changes have occurred but they have primarily been in abundances and size structures of resource species, trophic structures and the loss of local intra-specific populations, along with accompanying changes in genetic diversity. Those are all potentially-measurable facets of biodiversity but they are poorly, if at all, reflected by conventional indices of ecological diversity applied at the species level. That conclusion cannot be fully extended to coastal waters and certainly not to semi-enclosed marine areas, nor to most freshwaters. As focus moves from the open sea towards small ponds, the nature of biodiversity change in aquatic ecosystems becomes closer to what is seen in small terrestrial systems, though still not identical. One noted difference is that coastal extirpations, or at least those that have been recorded, tend to be of species at higher trophic levels, whereas invaders tend to be of lower level. The result is an enhanced trophic skew, even if overall species richness is unchanged (Byrnes *et al.* 2007; Stachowicz *et al.* 2007). Byrnes *et al.* (2007) found that 70% of the species listed by Dulvy *et al.* (2003) were either secondary consumers or higher-level predators. In contrast, amongst the species in four lists of introductions (for the Wadden Sea, San Francisco Bay, the Gulf of the Farallones and the Australian coastline generally), most were benthic suspension or deposit feeders – sessile planktivores being the single most common group.

2.5 A SPECTRUM OF DIVERSITIES

To summarize this Section 2, we suggest that the various concepts that have been given the "biodiversity" label (excluding the notion that synonymizes the term with "Life") can be viewed as a spectrum. At one end stand the trio of measures that were developed in an attempt to summarize Species Abundance Distributions as scalar values, *viz.*: richness, evenness and their combination, ecological diversity *sensu stricto*. At the other lies an idea that every desirable characteristic of an ecosystem can be considered a facet of an overarching "biodiversity" – though we here label that umbrella "environmental status". Somewhere between those extremes lies biodiversity as defined in the CBD. Biodiversity in that sense, which is the meaning attached to the term in this report, incorporates ecological diversity and all other facets of environmental status that concern the variability of living things. It is a subset of environmental status, while ecological diversity is a subset of biodiversity.

Vague though the concept of ecological diversity may be, by scientific standards, it is much closer to being definable than is environmental status. Ecological diversity is also quantifiable, albeit quantifiable in too many contrasting ways and at the cost of considerably more labour than is generally realized (see Section 4). Yet ecological diversity was originally conceived (if not exactly defined) as something that could be quantified by one simple index or at most a trio of indices. It is that simplicity which makes ecological diversity measurable but it necessarily limits the amount of meaning that it can carry. Richness, evenness and ecological diversity *sensu stricto* do bear some relationship to community productivity and stability but they are poor measures of either, productivity being linked to the diversity of functional groups, while stability relates to the diversity of redundant species, each of which is only weakly reflected by ecological diversity at the species level. In terrestrial and freshwater ecosystems, ecological diversity can sometimes be a useful surrogate measure for anthropogenic degradation, though that must be established on a case-specific basis. In the open sea, however, the effects of human pressures primarily appear as declines in abundances and shifts in size, trophic and population structures – none of which are meaningfully captured by the metrics of ecological diversity. Hence, those are all but useless when monitoring temporal trends in open-sea systems, though they may have some remaining value in mapping work.

Measures of ecological diversity are little more than highly simplified summaries of SADs – or of distributions of relative abundances of units other than species. Ricotta (2005) suggested that they be seen merely as summary statistics that carry no intrinsic meaning, any more than similarity indices do. We concur. Indeed, we see no scientific justification for the continued use of the metrics of ecological diversity and would prefer to recommend that they be discarded entirely. For reasons of widespread non-specialist perception and expectations, however, we fear that richness, evenness and ecological diversity *sensu stricto* must be retained, particularly where monitoring and routine reporting are concerned. Hence, we have to recommend that indices of ecological diversity be used within broader biodiversity-monitoring programs.

In taking that position, we recognize that others disagree. In particular, the BIP has not recommended ecological diversity indices for use under the CBD. The Convention's Conference of Parties has recently accepted an extensive list of "headline indicators" and "operational indicators" for monitoring of its Aichi Biodiversity Targets[11], amongst which only the monitoring of "genetic diversity of species" touches on the metrics of ecological diversity *sensu lato* and that only as regards genetic aspects. While not explicitly focused on biodiversity, OSPAR has developed "Ecological Quality Objectives" ("EcoQOs") that make no mention of what is here termed ecological diversity (see Heslenfeld & Enserink 2008), while ICES (2005) rejected ecological diversity indices for use in monitoring the ecosystem effects of fishing because of their lack of sensitivity to anthropogenic activity and the lag times between pressures and responses[12] – though ICES

---

[11] UNEP/CBD/COP/DEC/XI/3, December 2012.

[12] ICES (2005) blamed the deficiencies on "the currently available range of diversity measures as operation metrics". We would rather say that the failing, at least where marine ecosystems are concerned, lies in the concept of ecological diversity.

ICES (2005) further, and curiously, claimed that "diversity indices vary considerably from year to year". The indices do not, nor does ecological diversity often change swiftly in the sea, though sample diversities can be highly variable. If those were (incorrectly) substituted for community

(2005) recognized that monitoring would thereby fail to address a principal driver of their work: the commitments made by European governments to the CBD.

In contrast to the over-simplicity of ecological diversity, environmental status embraces complexity. In consequence, it is not quantifiable as an entire entity. Mayer (2006) and Sarkar (2010) emphasized that only selected facets can be given operational definitions and quantified. We will argue in Section 5 below that, while individual facets can be quantified by appropriately complex means, few can usefully be evaluated using simple indices. Further, the complexity of ecosystems is multidimensional. It will never be possible to reduce environmental statuses to a single, unidimensional measure, as some sort of "score". Different facets of an ecosystem will often have differing statuses, and even contrasting temporal trends, which must be understood if appropriate management decisions are to be made. Attempts to represent environmental status by a single number, let alone by a binary ("good" *vs.* "bad") score, will always be more misleading than helpful. Hence, measurement and monitoring of environmental status is, and will remain, challenging. The *UK Biodiversity Indicators in Your Pocket* series provides one illustration of how the task may be approached in practice. The Department of Fisheries and Oceans has recently explored others, with specific reference to the *Oceans Act* MPAs and AOIs (DFO 2010, 2011; Kenchington 2010, 2013a; Loseto *et al.* 2010). None are fully satisfactory, if only because broad-ranging environmental status must be quantified within the constraints of limited budgets. Both the Canadian and the British examples have unavoidably fallen back onto reliance on existing data streams, though those are unlikely to capture all critical facets of the status of many ecosystems.

Between the comparative rigour, but lack of meaning, of ecological diversity and the meaningful but nebulous environmental status lies biodiversity, in the sense of the CBD definition. That is important to conservation and management but it does not encompass everything about ecosystems that is important, being confined to the variability within them. Other facets of those systems merit attention as aspects of ecological status but they should not be confused with biodiversity, nor should biodiversity be mistaken for a summary of those facets. Biodiversity differs from ecological diversity in at least two ways. For one, the former encompasses a wider array of summaries of SADs, including ones that consider the community composition, in the sense of the named species present, and others that preserve the entire distribution of relative abundances instead of reducing it to a single number. Secondly, biodiversity explicitly reaches beyond species diversity to encompass the diversity of functional traits, phylogenetic diversity, intra-specific genetic diversity, variety in sizes of individual organisms and the diversities of habitats and landscapes. Indeed, it goes beyond such diversities of the components of biotic systems to also encompass diversity of physical structure (e.g. four-dimensional spatio-temporal arrangements) and of processes (Noss 1990). That far greater breadth allows biodiversity to capture

---

diversities, instability in index values would be expected, though that would not be the largest problem arising from the methodological error (see Section 4).

more meaning than ecological diversity can carry but it greatly complicates measurement. We will argue below that those complexities largely move biodiversity beyond the scope of the metrics considered in this report.

3 HILL NUMBERS AND OTHER THEORY

The development of diversity metrics and related theory has been in progress for upwards of half a century but continues unabated. Indeed, work has accelerated through the past two decades, in wake of the CBD. Some of the recent progress, especially Jost's (2006, 2008, 2010a, b) return to the ideas of Hill (1973), has transformed understanding of the metrics. Key theoretical issues are addressed here, as a transition between the conceptual material of Section 2 and the consideration of practical applications in Sections 4 and 5.

3.1 "TRUE" DIVERSITY

3.1.1 Hill Numbers Resurrected

Building on work by Renyi in 1961, Hill (1973) proposed a new way of quantifying ecological diversity *sensu stricto* using (in the symbols preferred by Jost 2006):

$$^{q}D = \left(\sum_{i}^{S} p_i^{\,q}\right)^{1/(1-q)}$$

where p_i is the relative abundance of the i^{th} of S species and q is termed the "order" of the diversity. It controls the sensitivity of ^{q}D to the presence of rare species in the community: ^{0}D accords all species equal weight, without regard to their relative abundance, while higher values of q accord less weight to rarities (Jost 2006). The sequence of ^{q}D for various q have become known as the "Hill Numbers".

Hill (1973) fully explained the virtues of his representation of diversity but failed to attract much attention. Perhaps his ideas came too late to be taken up by the first generation of quantitative ecologists. Pielou (1974, 1975, 1977), for example, noted some of the same points but only in connection with Renyi Numbers (the logarithms of Hill Numbers). In her later textbook, Magurran (1988) outlined many of Hill's (1973) ideas but her presentation of them was scattered and their significance lost. Heip *et al.* (1998) did provide a thorough account of the Hill Numbers, commending them as "a coherent system for diversity estimates", but in that era biodiversity was a matter for conservation biology and of limited interest in quantitative ecology. It therefore fell to Jost (2006) to return to Hill's (1973) work, promoting it with renewed force and thereby re-orienting concepts of what ecological diversity is. In doing so, Jost (2006) presented the material as though much of it was novel but most of his technical content merely repeated Hill's (1973) explanation of three decades before. Thus, Jost's (2006) contribution lay in the force, clarity and perhaps timing of his arguments. We follow his presentation of them here.

Jost (2006) began by distinguishing "diversity" itself (by which he meant what is here termed ecological diversity *sensu stricto*) from the indices which have been used to quantify it, arguing that the former exists independent of deficiencies in attempts at its measurement. He also stressed that indices need not be linearly related to what they are supposed to be measuring – a point that he illustrated by noting that the radius of a sphere is an index of its volume but attempts to substitute the radius for the volume in engineering calculations would lead to major errors.

The familiar Shannon index (*H'*) was originally developed as a measure of the entropy of a code, for use in information theory, and many other indices of ecological diversity have subsequently been shown to be entropy measures. When applied as an index of species diversity, *H'* measures the uncertainty in which species would be represented by the next specimen drawn at random from a community. Jost (2006) commended the Shannon index, calling it "the most profound and useful of all diversity indices", but uncertainty over what species might be picked blindly is obviously not synonymous with what most ecologists would understand by the term "species diversity" – which was Hurlbert's (1971) reason for dismissing *H'*[13]. For the latter, Jost (2006) suggested that a community with sixteen equally-abundant species should have twice the diversity of a community with eight species, again equally abundant. The values of *H'* of two such communities would, however, differ in the ratio $x+\log_b 2 : x$ (meaning 4:3 if the logarithmic base used was 10, approximately 5:3 if the base was *e*). Hence, *H'* may be a useful index of ecological diversity but it cannot be diversity.

Jost (2006) proposed that, far from being anything to do with entropy, diversity itself is proportional to the number of species in a hypothetical community which shares the same value of a diversity index as the community being evaluated and in which every species is equally abundant. By setting the constant of proportionality at unity, diversity can be measured as a number of species; not the species richness of a real community but its "effective number of species", meaning the number in the hypothetical community with equal abundances. That idea had been foreshadowed in quantitative ecology by MacArthur in the 1960s and it has long been familiar in other sciences which utilize entropy measures. Clearly, quantifying diversity in units of "effective species" has an intuitive appeal lacking when the measurement is in units of entropy (Jost 2006). The key point in

---

[13] Hurlbert's (1971) seems an unreasonably pedantic position, founded in the mistaken notion that ecological diversity should have a mechanistic relationship to the structure and function of ecosystems – mistaken because the structures of even the simplest of ecosystems are vastly more complex than can be captured by any index. As Pielou (1975) wrote in promoting the use of *H'*: "it cannot be too strongly emphasized that fancied links between the information-theoretic concept of 'information' and the diversity of an ecological community are merely fancies and nothing more". Her arguments for the use of *H'* did not invoke information content at all.

Hill's (1973) and Jost's (2006) argument is that this "effective number of species" can be quantified using the Hill Numbers.

It is also important that, for some values of *q*, ${}^{q}D$ is a simple transformation of a familiar diversity metric – Jost (2007) later terming ${}^{q}D$ the "numbers equivalents" of the older indices. With $q = 0$, for example, ${}^{0}D$ reduces to simply *S*, the species richness. Simpson's index of concentration:

$$D = \sum_{i=1}^{S} p_i^2$$

varies inversely with diversity but adaptations of it have long been used as indices of ecological diversity nonetheless, most commonly in the form of the familiar Gini-Simpson index (1-*D*) but alternatively as the Inverse Simpson index (1/*D*) and occasionally as –log *D* (Pielou 1974, 1975). ${}^{2}D$, the "effective number of species" of order 2, is simply 1/*D*, the Inverse Simpson index. With $q = 2$, it moderately down-weights the influence of rare species. A less-commonly encountered metric is the Berger-Parker dominance index, which is the relative abundance of the single most abundant species in the community of interest. ${}^{\infty}D$ is equal to the inverse of the Berger-Parker index. It entirely ignores not only all rarities in the community but also all abundant species, save the single most abundant.

Those higher orders of ${}^{q}D$ have their uses, as the equivalent familiar indices have long done, but the order of broadest utility, ${}^{1}D$, is that which weights all species by their relative abundances. The exponent of the ${}^{q}D$ equation is undefined when *q* is unity but the limit of the equation is then equal to the exponential of the Shannon index, *H'*, which is to say it equals $e^{H'}$. Jost (2006) praised that measure of diversity for "its unique ability to weigh [*sic*] elements precisely by their frequency", but he was at pains to insist that it is derived from the concept of effective species and the series of Hill Numbers, not from the entropy of information theory. He rejected any suggestion that it should be preferred simply for reasons of prejudice or fashion – though it does have the great advantage of broad familiarity.

As a family of indices, ${}^{q}D$ passes a critical test: Given a community of *S* species, each with some relative abundance of its own, divide each species into two units, such that each new unit has one half the relative abundance of the species from which it came. The resulting new community has 2 *S* kinds and yet, in Jost's (2006) view, identical evenness to the original community (though the validity of that contention will be taken up below: see Section 4.4). Hence, many ecologists would agree that the new community has double the diversity of the old, yet few of the established diversity indices would double their values when confronted with those data. Jost (2008) later offered two examples of such failures: First, the Gini-Simpson index would return a value of 0.999999 for the diversity across a continent containing a million equally-abundant species but 0.99 if there were just 100 species – a reduction of less than 1% in the index value when the community lost 99.99% of its diversity. The second example was more realistic: A field study of the trees on Barro Colorado Island, Panama has found that the Gini-Simpson index yields a value of 0.95 for their diversity. If there were 20 such islands, each with the exact same relative abundances of their species but with no species in common, the index would return a value of 0.998 for the combined set of islands. Taking those numbers on their face, it would be a reasonable supposition that 95% of the tree diversity could be protected by preserving just one of the 20 islands, while clear-felling the rest, when that strategy would obviously result in conserving just 5% of the tree species originally present (Jost 2008). Clearly, the non-linearities between the numerical values of the common diversity indices and the perceived diversities of communities are so severe that the numbers risk being seriously misleading, instead of informative. In contrast, ${}^qD$ does pass the test, doubling in value when the number of species is doubled while preserving relative abundances, and does so for any values of q, S or the relative abundances. That property leads ${}^qD$ to perform as one would intuitively expect diversity to do when confronted with various features of real datasets (Jost 2006).

3.1.2 True Diversity Debates

Thus far into his argument, Jost (2006) made a very good case for preferring ${}^qD$ measures. However, what gave his presentation its power, yet at the same time provoked sharp debate, was that part way through his paper he chose to confine the term "diversity" to his ${}^qD$. To avoid confusion with what he and others had previously used that word for, he then took to calling ${}^qD$ "true diversity", casting all other measures as mere indices of ${}^qD$. Jost's (2006) choice of terminology implied that most of the authors who had gone before had addressed, to some degree, various "false" diversities (cf. Gorelick 2011), which may explain the resistance that his proposals have met[14].

In one notable response, Hoffmann and Hoffman (2008) noted that what they termed "biological diversity" is too complex a concept to be captured by a single metric. It is but that was only a sterile semantic point, arising from a failure to understand that the one term, "diversity", was being used for the two very different things that are here labelled ecological diversity and environmental status. Hoffmann and Hoffman (2008) further suggested that what diversity is should be a matter of "desirable properties", which have to be examined in relation to a particular context. It is then necessary to find a metric that matches the properties of concern. They disputed that ${}^qD$ is diversity, considering it as merely a mathematical expression[15]. They also noted that perceptions of species diversity are not simply linked to the number of species, since a community of S species all drawn from the same genus or family would appear less diverse than one of S species that contained representatives of a number of phyla. In that, however, they moved into the field of phylogenetic diversity, setting themselves against the long-established intent of metrics of ecological diversity ["species neutral" measures, as Jost (2009a) termed them in his response], which has exactly the limitation that they identified. More tellingly, Hoffmann and Hoffman (2008) questioned the relevance of ${}^qD$ to genetic, functional or ecosystem diversities, while insisting that quite other functions would meet Jost's (2006) criterion for a "true diversity", some of which would not track diversity at all. By implication, there is nothing absolute or uniquely "true" in ${}^qD$. Hoffmann and Hoffman (2008) conceded that the latter may be a very reasonable measure for many purposes but denied that there is any "correct" definition of diversity, only useful ones – which was as good a way as any of saying that their "diversity" is not a scientific concept at all [though their position was consistent with those of Mayer (2006) and Sarkar (2010) noted in Section 2 above].

---

[14] Tuomisto (2011) has suggested that "true diversity" should be understood as equivalent to the way that taxonomists use "true" to attach a common vernacular term to a particular taxon. Jost (2006) himself made no such suggestion, however and Tuomisto (2012) has since written of ${}^qD$ as "diversity itself".

Despite maintaining the "true" label in papers published as late as 2009 (Jost 2007, 2009a), Jost avoided using it in two recent contributions (Chao *et al.* 2010; Jost *et al.* 2011), except as necessary to link ${}^qD$ to what he had previously called "true diversity".

In his reply, Jost (2009a) repeated his earlier arguments, clarifying parts, but he also drew the sensible point that when ecologists interpret the values generated by diversity indices, they tend to suppose that the index values are proportionate to diversity, while it is clear that there are major non-linearities between the commonly used indices and what those same ecologists mean by "diversity". While true, that only supported Hoffman and Hoffman's (2008) point that the meaning of "diversity" is something to be set out in prose, while measures of it must be selected to match that meaning as best can be. Seen thus, ${}^qD$ is a useful measure of ecological diversity (in so far as ecological diversity has any utility at all) but it is not ecological diversity itself. ${}^qD$ is certainly not biological diversity, as a whole, nor an index of more than one facet of it.

Gorelick (2011) took a rather different route in his criticism of Jost's (2006) proposals, accepting as a starting point that any index of ecological diversity only captures what he termed "certain nuances" of an ecosystem, leaving aside much else that is not of immediate concern – for the very good reason that, if other issues were of concern, the ecologist should not be using a diversity index at all. Even within what such indices can capture, there are a variety of "nuances" and Gorelick (2011) recommended a pluralistic approach. He, indeed, raised one of the obvious defects in Jost's (2006) position: the latter's "true" diversity is actually an infinite variety of alternative diversities, with q being freely set anywhere $0 \leq q \leq \infty$. It is difficult to see how all of those alternatives can be simultaneously "true". Another weakness that Gorelick (2011) identified was that not all established diversity indices can be converted into Hill Numbers, so that some indices which have had proven value are excluded from Jost's (2006) system. He suggested replacing the inflammatory term "true diversity" with "effective species diversity" or something else less provocative. Gorelick (2011) favoured the Shannon index in its original form (not as the exponential) and also the square of the coefficient of variation, used as a diversity index (which is one that cannot be converted into a Hill number), since each of those can be related to the observed-over-expected ratios of χ^2 tests. His conclusion was that indices of ecological diversity should be chosen based on the ecological question being asked, without raising any particular index to the status of being "true", thus maintaining a plurality of alternatives (Gorelick 2011). A response to those points was published by Tuomisto (2011), who has emerged as one of the principal proponents of Jost's (2006) vision, but she barely rose above semantic issues revolving around the use of the "true" label.

---

[15] Showing just how confused efforts to untangle these concepts can be, Tuomisto (2010c) insisted that "Diversity is simply a variable that can be measured for a dataset" and distinguished that from "the factors that drive species diversity", the latter presumably being more or less what Hoffman and Hoffman (2008) saw as "diversity", distinct from its mathematical approximation.

A second obvious deficiency in Jost's (2006) argument is that ${}^0D$ is species richness and hence not ecological diversity *sensu stricto* at all, since that is defined to be a combination of richness and evenness and cannot be equal to either of those alone (cf. Moreno & Rodríguez 2011). Similarly, ${}^\infty D$ is a measure of evenness that ignores richness and so is not ecological diversity *sensu stricto* either[16]. One might perhaps argue that richness is a measure of diversity that gives zero weight to evenness (i.e. $q = 0$) and that evenness is a measure of diversity that gives zero weight to richness (i.e. $q = {}^1/_0 \Rightarrow \infty$) but that would be a semantic stretch. It seems better to accept that Jost's (2006) "true" diversity includes richness and evenness, which are not themselves ecological diversity *sensu stricto*, true or otherwise.

Tuomisto (2010a) argued for "true" diversity because (1) it is the only measure of diversity that doubles when intuitive understanding of diversity does, (2) it can be logically interpreted as the "effective number of species" and (3) it allows adjustment of sensitivity to rare species by varying q. The second advantage is real: reporting to a lay audience in terms of numbers of species is apt to be more understandable than if an abstract number expressed to two decimal places is stated, though that same audience may struggle to comprehend the difference between species richness and the "effective" number. The third supposed advantage is not one: the same flexibility was always available by changing the index used (from richness to Shannon and on to Gini-Simpson, for example), while insisting on the "true" status of ${}^qD$ regardless of the value of q never aided clarity. Rather, it is the proportionality between ${}^qD$ and intuitive understanding which gives the idea is real point, particularly when reporting to a lay audience – which is a focus of the present report. Indeed, in more recent contributions, Jost himself has stressed that characteristic (e.g. Chao *et al.* 2010).

---

[16] Some writers (e.g. Magurran 2004) have confused metrics of ecological diversity *sensu stricto* that are relatively insensitive to rarities, such as the Gini-Simpson and Inverse Simpson, with metrics of evenness and dominance. ${}^\infty D$ is a measure of evenness, as ${}^0D$ is a measure of richness, but ${}^2D$ is an index of ecological diversity *sensu stricto*.

3.1.3 Application and Recommendations

For our immediate purposes, it is fortunately possible to agree with all sides in this tortuous debate. Ecological research should perhaps not use indices of ecological diversity at all, since there are much more informative ways of summarizing the information in a list of species abundances. If a research scientist nevertheless sees value in a particular index for a particular purpose, he or she should always have the opportunity to use that metric, justifying the choice as every aspect of a study's methodology must be justified. In that context, Gorelick's (2011) pluralism is surely appropriate. In monitoring or mapping work in support of management, however, such flexibility can only lead to confusion, since it would prevent inter-comparability while confusing the non-specialists who are the intended end-users of the results. Biodiversity, in the CBD sense, is a far wider topic than any one index can capture and its effective monitoring needs a broad program tracking multiple indicators. To the extent that some index of ecological diversity should be included amongst those indicators, however, it would be appropriate to standardize across the Department's programs.

When deciding which to select, it is necessary to consider what "desirable properties", in Hoffmann and Hoffman's (2008) sense, the standard index should have. The foremost one, we suggest, is that its meaning should be readily understood by non-specialists (an aspect of "concreteness", *sensu* Rice & Rochet 2005), which implies that variations in the numerical value of the index should be broadly proportionate to what end-users perceive to be differences in the diversity of communities: a loss of one tenth of the species in an area, while evenness is maintained constant, for example, should result in 10% decline in the index value. That is a property of the ${}^{q}D$ series but is rarely, if ever, seen with other known indices of diversity. The ability to report ${}^{q}D$ in understandable units of "effective species" is a further advantage.

Species richness, $S = {}^{0}D$, is simple and understandable but discards much expensively-collected abundance information. It is also almost completely insensitive to anthropogenic change in the open sea, as explained above (Section 2.4). Hence, we cannot recommend ${}^{0}D$ as the primary measure of ecological diversity for routine use. Where they can be, its values should be determined and reported nonetheless, not because of the arguments of Jost (2006) but because at least two of richness, evenness and ecological diversity *sensu stricto* are required if any of them are to be meaningfully interpreted, while presenting all three adds clarity with minimal extra computation (see Section 2.1.1). However, despite frequent claims to the contrary, determining species richness requires considerably more data than are needed to quantify ecological diversity *sensu stricto* (see Section 4.2) and in some cases it may be necessary to forego richness (and thus evenness also).

In the past, some authors (e.g. Magurran 2004) have preferred indices which transform into ${}^{q}D$ with $q > 1$ (such as the Gini-Simpson), on the grounds that they converge towards final values more quickly than does the Shannon index, thus providing firm results with fewer samples. Yet, in light of Jost's (2006) work, that swifter convergence can be seen to arise from the lesser weight given to rare species when $q > 1$ – the lesser weight meaning that the extensive sampling required to document rarities can be evaded without much loss of precision (Jost 2007). Where the objective is research into ecosystem structure and function, with an emphasis on the abundant species that are responsible for most of the production and consumption, that might well be a sensible alternative – if any metric of ecological diversity was thought to be a useful measure of structure or function. In the case of biodiversity monitoring or mapping, however, down-weighting of rarities would be counter-productive since their conservation is a special concern. Rice (2003) argued that particular indices of ecological diversity *sensu stricto* should be chosen such that they would be sensitive to the kinds of change expected and hence to the primary anthropogenic pressures of concern. Where fishing dominates, changes in evenness, rather than richness, are to be expected and hence Rice (2003) called for what he termed "evenness-weighted" indices, which is to say those with higher q. Where the concern is pollutants, "richness-weighted" indices (those with low q) would be preferable. That idea has much merit but a standard approach applicable across all Canadian aquatic ecosystems would not be advanced by encouraging local adaptations, while we recommend reporting richness and evenness themselves, in addition to some index of ecological diversity *sensu stricto*. Hence, we do not recommend ^qD with $q > 1$ for general application in the Department's monitoring and mapping.

Magurran (1988) reviewed some early examinations of the ability of various indices of ecological diversity to distinguish among samples from (terrestrial) communities with small differences in species compositions, which studies suggested greater discriminating power within the ^qD family when q is low – meaning that the communities differed more in the number of rarities present than in the relative abundances of the commoner species. In one study, maximum discrimination occurred at $q \approx 0.5$. While further investigation of that option would be worthwhile, the expectation that anthropogenic change in the sea should be reflected in evenness, rather than richness, argues for a higher q. Moreover, integer values of q that allow ^qD to be a simple transformation of a familiar index would have considerable benefits when communicating results.

Combining all of those considerations, we recommend that the appropriate standard index to use in reporting diversity monitoring is ^1D, which is to say $e^{H'}$, the exponential of Shannon's H'. That has the benefit of being closely related to the well-known and oft-calculated Shannon index. Indeed, the "exponential Shannon" index, often symbolized as or EXP H', was itself recommended by Peet (1974) nearly four decades ago and has been increasingly used in recent years, irrespective of Jost's (2006) work. Jost (2007) and Jost *et al.* (2011) noted the unique properties of ^1D, while the latter foresaw it becoming the standard measure of ecological diversity. We concur.

It is often said that the Shannon index is "insensitive" (cf. Magurran 1988). The primary cause lies in its curvilinearity with respect to perceived "diversity". That deficiency is resolved by the use of $e^{H'}$. It remains true that EXP H' is less sensitive to changes in the number of rare species than is richness, S, but that is a deliberate intent of metrics of ecological diversity *sensu stricto* and a reason to report them alongside values of S, not an argument for discarding either metric. Other objections to the Shannon index were reviewed long ago by Washington (1984) and can still be encountered, often in garbled form, in more recent literature. One class of them objects to the metric because it has no direct biological interpretation. While that may be true (though, following Jost, values of the exponential Shannon metric can be read as numbers of "effective species"), the concocted interpretations sometimes offered for other metrics of ecological diversity are largely spurious. All such measures are merely summaries of SADs and should be judged as such (cf. Pielou 1975) – the exponential of the Shannon index being the best alternative for many purposes. A second class of objections centres on values of the metric being generally uninformative. They most certainly are – but that is a deficiency of ecological diversity, inevitably shared by all metrics of that concept, not a fault of the Shannon index alone. Much more recently, Magurran (2004) saw no merit in either H' or ^1D, though she did not raise any objections other than those which Jost (2006) himself directed at H' and resolved by transforming it into ^1D. In his turn, Jost (2007) dismissed Magurran's (2004) arguments in favour of richness and the Gini-Simpson index[17].

3.2 DIFFERENTIATION OR β DIVERSITY

In research work, it might be meaningful to compare the ecological diversities in, for example, several eelgrass beds. It might be meaningful to compare the average diversity in those beds to that in similarly-sized areas of open sand bottom. A comparison among multiple plots, some of which lie in eelgrass, others on sand and yet others spanning across the two habitat types could, however, be misleading, since the arbitrary choice of boundaries in a study design would result in inflated values for the measured diversity in plots that contained both eelgrass and open sand. Should the choice of boundaries not be arbitrary, meaningful questions might nevertheless be asked of the data, such as whether some bays have higher species diversity than others because they have higher habitat diversity. In the sea, all species have distributions that are to some degree depth-related and for many the effects of depth are extremely important. Study plots that have relatively homogeneous depths, such as those on the flat tops of offshore banks, may thus be expected to show lower ecological diversity than similarly-sized plots that span wider ranges of depths, not because any ecological communities in the latter are more diverse but because arbitrarily-selected sloping areas include a wider variety of depths, and hence of habitats and species, within the same number of square kilometres. If the aim of a study was to select the smallest MPA that contains examples of the most species, the different measured diversities of flat and sloping areas may be relevant but if the structure and function of the ecosystems are of interest, it would be erroneous to suppose that communities on slopes are more diverse.

---

[17] McGill (2011) took a quite different position. He rejected what we term ecological diversity *sensu stricto* in favour of reporting richness and evenness. When considering directly comparable communities (which would usually include looking for temporal change in the same spatial community) with constant sample sizes, he recommended reporting S and either J' or else a parameter of a log-logit probability distribution. To those, he suggested adding a dominance measure (e.g. McNaughton dominance) and a rarity measure (e.g. log skewness). With markedly unequal sample sizes, McGill (2011) saw the deficiencies in direct comparisons of sample richness (though he did not discuss the long-established alternatives for avoiding those) and claimed similar difficulties with abundance measures. Oddly, he also thought that those would be problematic when comparing dissimilar communities. For both situations, he recommended replacing N with the c parameter of the logseries and S with Margalef's index – a crude estimator of asymptotic S. Our reasons for rejecting those suggestions are stated elsewhere in this report.

In a series of papers and books published between the 1950s and 1970s, R.H. Whittaker sought to formalize these complexities. He coined the term "α diversity" for the ecological diversity within an environmentally homogeneous patch[18], while that within a larger "landscape" (containing multiple, heterogeneous patches) he designated as "γ diversity". The ecological diversity within an entire region or biogeographic zone, containing a variety of landscapes, can be labelled "ε diversity" (Magurran 1988). Some authors (e.g. Gray 2000) have recommended using rigorous definitions of the hierarchical levels to which α, γ and ε diversities should be applied. In practice, however, truly discrete, homogeneous patches are unusual. Real ecosystems are spatially heterogeneous, comprised of patches within patches, the boundaries between being clinal and the whole overlain on larger-scale clines. Moreover, the distinction between a "landscape" and a "region" must always be an arbitrary one. Hence, ε diversity is rarely invoked. Instead, α diversity is taken to be the ecological diversity within one example of the finest subdivision of the ecosystem that is deemed relevant to the purposes at hand. γ diversity is then the ecological diversity within some broader area, also relevant to the questions of immediate interest. The γ diversity of some area is never less than, and usually very much higher than, the a diversities in each of the patches within that area.

If a particular research study must examine more than two levels, the γ of the lowest-level comparison can be treated as the α' of the next one, γ' as α'' and so on (cf. Tuomisto 2010a). Thus equipped, the distribution of ecological diversity throughout any ecological hierarchy can be examined, without need to invoke more than the first and third letters of the Greek alphabet, while responsibility for determining which two or more hierarchical levels are relevant to the questions at hand lies with the designer of the study – where it belongs.

---

[18] The term "α diversity" is sometimes equated to richness (e.g. Swingland 2001). That seems unwarranted.

Whittaker's approach to these issues has become the accepted one but others are possible. Pielou (1975), for one, considered mosaic distributions from a perspective of the inter-relationships among particular patches, rather than simply quantifying patchiness. She only invoked the term "β diversity" in relation to the rate of species turnover along a gradient.

It is sometimes recognized that α diversity is specifically the ecological diversity of a "community", even one arbitrarily delimited, and not that of the organisms taken in a single sample, though other authors do not accept the distinction (e.g. Tuomisto 2010a). The diversity within a sample may best be termed "sample diversity", though Whittaker's term was "point diversity" (Magurran 1988) – the latter being misleading when the sampling gear is worked over an appreciable distance, as with a trawl or dredge, and thus may take specimens from a number of diverse patches. Sample diversity may be useful for some purposes but it has less ecological meaning than the α diversity of a community.

It is unlikely that monitoring or mapping work in support of management will need to be concerned with more than one level in an ecological hierarchy and the diversities addressed by the conclusions of this report are all γ diversities. They are, however, rarely the same as the diversities of available field samples (considered as α diversities though technically sample diversities). As will be explained in this Section and in Section 4.2.5, a failure to appreciate that difference leads to one of the commonest mistakes in applied diversity measurement – a mistake that has led to erroneous conclusions in recent work within DFO (see Section 4.2.5). Hence, it is essential that both diversity analysts and those who use the diversity values have a working understanding of modern applications of Whittaker's ideas.

He grouped sample, α, γ and ε diversities as four levels of "inventory diversity", distinguishing them from "differentiation diversity", which relates to the differences among the smaller ecological units within a larger unit, rather than to the species diversity within any unit. Differentiation diversity is thus a way of considering habitat diversity (an important consideration within biodiversity) in terms of the species diversities of the habitats, though it can also be seen as a measure of the variability in species diversity. In theory, the differences among homogenous patches within a landscape are termed the "β diversity", while those among the landscapes within a region are the "δ diversity" (Magurran 1988). The latter term has been as little used as ε diversity but β diversity, as the step between (broadly defined) α and γ diversities, has been much invoked.

Unfortunately, Whittaker himself offered a considerable number of (mutually contradictory) quantitative meanings of his β diversity, while a good many more have been added by subsequent authors, all of them seeking to measure the additional variability in species abundances that arises from broadening the variability in habitats encompassed by a study. The result has been a morass of confusions, leading to values for β diversity that are in no sense inter-comparable (Jurasinski *et al.* 2009; Tuomisto 2010a, b). One source of the complications is that the inventory diversities all concern the diversity of species (or other units), albeit across different spatial scales. The differentiation diversities, in contrast, are diversities of spatial or quasi-spatial units, be they sampling units, homogeneous communities or "landscapes" (Tuomisto 2010a). α, γ and ε diversities are thus variations on a theme, while β and δ diversities are variations on a different theme and have to be considered separately from the inventory diversities – a complication too easily missed in introductory textbooks. It has been argued that differentiation diversity is not properly "diversity" at all and that that term should be restricted to inventory diversities. However, while some of Whittaker's variant suggestions for β diversity do not bear any close resemblance to ecological diversity, Tuomisto (2010a) has shown that both the differentiation and inventory forms can be measured using equivalent units and hence, when appropriately quantified, can be regarded as variants of the same underlying concept – though still not the same as one another.

Jurasinski *et al.* (2009) recently reviewed the many alternative ways of quantifying β diversity, providing an accessible summary of the complications but little in the way of a resolution. Their work was followed by an extensive review prepared by Tuomisto (2010a, b) in which she placed β diversity within the context of Jost's (2006) "true" diversity and erected a structure into which she fitted some (but not all) of the previous alternative quantifications – though it cannot be said that her work promoted clarity of understanding among her readership. Her overall conclusion (Tuomisto 2010b) appears to be that the many different ways of representing β diversity measure genuinely different things and hence are not inter-comparable, from which it follows that authors must be clear about what measures they used and careful about correctly interpreting their results – unremarkable recommendations but ones which, nonetheless, seem to have been forgotten in too many past studies.

Long before, Pielou (1974, 1975) had preferred the Shannon index over all other metrics of ecological diversity, coming close to rejecting the rest entirely, because of its unique additive property: for example, in the case of bottom-trawl surveys conducted at various seasons through the year, it would be possible to determine the overall diversity of trawl-vulnerable groundfish species in the surveyed area, the diversity within any one season or the seasonal diversity for any one species (the latter perhaps distinguishing residents from migrants). If diversity is measured using the Shannon index, and only if that is the metric used, overall diversity is equal to the sum of the species diversity and the mean of the within-species seasonal diversities, which also equals the across-species seasonal diversity plus the mean of the within-season species diversities (Pielou 1974, 1975). Washington (1984), not unreasonably, objected to Pielou's argument on the grounds that she was more interested in the mathematical tractability of the indices than in their biological meaning but Pielou, more than many other early specialists in diversity measurement, was aware of the limited meaning of the metrics, correctly seeing them as summaries of SADs rather than as profound measures of ecosystem structure or function. Having rejected most biological meaning, she was left to select indices by their tractability.

Studies using the sorts of orthogonal comparisons of diversities that Pielou (1974, 1975) presented have been unusual, especially when compared to the total number of reports of diversity values, but the additivity of the values of the Shannon index have more utility when β diversity is under consideration, since the value of γ diversity is the sum of the values of α and β diversities if and only if they are all measured using the Shannon index. Unfortunately, that restriction appears to have been too easily forgotten and some authors have attempted to use the additive relationship between α, β and γ diversities when those are measured using other indices, such as the Gini-Simpson (Jost 2007).

Jost (2006) found merit in his ${}^{1}D$, which is to say the exponential Shannon index, $e^{H'}$, parallel to that which Pielou (1974, 1975) had found in H' because the additivity of the latter amounts to a multiplicative relationship of $e^{H'}$: when all are measured as ${}^{1}D$, γ diversity is the product of the values of α diversity and β diversity. Jost (2007) went on to demonstrate that, under certain circumstances, the same multiplicative relationship also holds true for all ${}^{q}D$, with $q \neq 1$ – the circumstances including all in which the various communities are given equal weights when calculating γ diversity, which is common in practical applications (Jost 2007)[19].

Tuomisto (2010a) took up that theme. As explained above (Section 3.1), inventory diversities can be quantified in “effective numbers of species”. Tuomisto (2010a) proposed that differentiation diversities should likewise be measured in “Compositional Units” (“CUs”), each containing an “effective number of species” equal to the average “true” α diversity at the lower of the two levels in the ecological hierarchy under consideration – each of those “effective species” being unique to the CU in which they belong. Tuomisto (2010a) showed that the number of CU is the “true” β diversity, in the same sense that the number of effective species is the “true” α or γ diversity. Again echoing Jost (2006), the number of CU is equal to the ratio of the γ to the α diversities and hence is a measure of how many times more diverse a whole dataset is compared to the average of the units within it. Tuomisto (2010a, b) went on to explore multiple complications at great length but those can be passed over here. Her central point is that Jost’s (2006) preference for relying on ${}^{q}D$ as the proper measure of ecological diversity can be extended beyond inventory diversities and applied to differentiation diversity, which is to say β diversity, with the unit of measurement being the CU. γ diversity is then fully decomposable into its α and β components. That advance has been strongly welcomed by Jurasinski and Koch (2011) but not by all other commentators. Gorelick (2011) for one has worried that Jost’s

---

[19] In keeping with some of his other recent papers, Jost’s contribution on β diversity (Jost 2007) provoked a published discussion (e.g. Ellison 2010; Jost 2010b; Veech & Crist 2010a,b). The points at issue do not, however, appear to have dealt with the substance of Jost’s (2007) argument but only with the nature and degree of independence between α, β and γ diversities. As Ellison (2010) noted, in empirical practice, α diversity is measured (or, often, estimated), γ diversity is estimated from multiple α values, after which β diversity is determined from the estimates of α and γ. Statistical independence cannot be achieved, though Jost (2010b) argued that orthogonality not only can but must be, while Veech and Crist (2010b) responded that it cannot be.

(2006) and Tuomisto's (2010a, b) choice of measures may conceal a reality that γ diversity is not fully decomposable and hence will discourage investigations into why it is not. It appears to us, however, that the "why" was only ever because of defects in the measurement scales formerly employed.

Tuomisto (2010a, b) offered little by way of practical guidance in the use of her ideas and it is unsure whether she has attempted to apply them herself. It is clear that the analyst could choose any value of q between zero and infinity but the results will not be meaningful unless the same value is used in the calculations of α, β and γ diversities under consideration. More seriously, all of Tuomisto's (2010a, b) theorizing concerned the extraction of values for the diversity in a dataset. Since complete censuses of chosen taxocenes within a community are very unusual, the available data will usually comprise a small subsample of all that could have been gathered. Extrapolating inventory diversities from samples to whole systems is a challenge that has faced ecologists for decades and Tuomisto (2010b) offered no new insights. She did, however, note that the calculation of β diversity from α and γ diversities makes the former subject to the biases between sample and community diversities in each of the latter. Those biases normally differ between the α and γ levels, which severely complicates extrapolating β diversity from data to ecosystem. Tuomisto (2010b) offered no broadly-applicable solutions and, pending further conceptual advances, it seems best to treat β diversity as something specific to a particular research program and hence to a particular dataset, without direct, quantitative application to real ecosystems.

Differentiation diversity at the species level may remain primarily something utilized, when appropriate, in research work but it does have some applications. Wiersma and Urban (2005), for example, have demonstrated its importance when planning networks of protected areas with "representation" objectives: if most of the β diversity in a region is expressed across moderate spatial scales, a single, large MPA could enclose examples of a majority of the community types found within that region. However, if the β diversity appears at larger spatial scales, multiple, smaller MPAs would be needed to achieve the same representation goals. Potentially, β diversity might serve as a valuable measure of habitat diversity based on the species that live in different areas, avoiding the need for a (partially arbitrary) classification of habitat "types" that could then be summarized using the indices of ecological diversity. That approach would, however, need to be examined through research applications before it could be recommended for routine application.

Perhaps of most importance, end users should remain aware of the varied ideas that have been given the label "β diversity" (cf. Jurasinski *et al.* 2009: Tuomisto 2010a, b). For example, Shackell *et al.* (2012) recently determined what they described as "β diversity" of large-bodied groundfish from Georges Bank to Sydney Bight, though what they calculated was the rate of change in a multivariate similarity index. That was an appropriate measure for the purposes to which Shackell *et al.* (2012) applied it and was also one of Whittaker's approaches to "β diversity" (indeed one most commonly encountered in recent literature: Jurasinski *et al.* 2009). Hence, its identification under the "β diversity" label cannot be faulted. However, rate of change in similarity is not β diversity in Tuomisto's (2010a, b) sense and is not closely related to ecological diversity at all.

3.3 GENETIC DIVERSITY

Data on allele frequencies can be reduced to indices analogous to those of ecological diversity. There are some important differences, however. One is that the factors which shape intra-specific genetic diversity (e.g. mutation, gene transfer, migration between populations, genetic drift, population abundances and the number of populations) are definable and reducible to mathematical models, with many of the model parameters being estimable. Hence, measures of genetic diversity can have a mechanistic foundation and are often interpreted in terms of particular mechanisms, such as when estimating rates of migration between populations (cf. Meirmans & Hedrick 2011). In contrast, and despite the many attempts at constructing ecological theories that purport to explain variations in species diversity (cf. Magurran 2004; McGill 2010; Rosindell *et al.* 2011), ecological diversity is shaped by multiple poorly-understood interactions among many different species, while measurements of that diversity are additionally shaped by the selective processes of data collection. Hence, metrics of ecological diversity can only be regarded as descriptive summaries of Species Abundance Distributions.

Conversely, when considering species diversity, the species themselves are of conservation interest. Whitlock (2011) has suggested that population genetics is different in that the alleles themselves are of no particular interest, being merely means by which to study demographic issues, while the loci selected for study are chosen because they are highly variable and hence are unrepresentative of the rest of a species' genome. While that was insightful, it is the amount of allelic and genotypic variation that is of interest when monitoring genetic diversity, just as the amount of variation in the species represented by the individuals in a community, not the particular species present, is the focus of species diversity.

Thirdly, whereas the metrics of ecological diversity began with the work of quantitative ecologists and the indices of most other aspects of biodiversity have been built on that foundation, quantification of genetic diversity is founded in the discipline of population genetics. There are parallels between the two, and even some cross-linkages, but also pronounced divergences. To add to the complexities, the two scientific communities have adopted contrasting symbologies, sometimes using the same symbol for quite different things[20].

---

[20] We have here maintained the original symbology of each discipline, on the expectation that that will cause less confusion than would any attempt at standardization, which would necessarily place us in conflict with the established usage of either ecologists or population geneticists.

Within-population genetic diversity at a single locus could be represented by the number of different alleles found in the population (analogous to species richness) or by the relative frequencies of the alleles among the individual organisms in the population (analogous to a Species Abundance Distribution). Any of the indices of ecological diversity *sensu stricto* could be applied to the latter values. In practice, however, the normal measure used is "heterozygosity" which, in a randomly-mating population, is equal to Nei's "gene diversity":

$$H = 1 - \sum_{i=1}^{k} p_i^2$$

where p_i is the relative frequency of the i^{th} of k alleles found at the locus[21]. That expression is algebraically identical to the Gini-Simpson index of ecological diversity. Given a structure of partially-isolated populations, Nei (1973) proposed a measure of differentiation among the populations:

$$D_{ST} = H_T - H_S$$

where the subscript S indicates the mean of the values for each of the populations[22] and T indicates the value for the pooled populations. To obtain a measure that could ostensibly vary between zero and unity, Nei (1973) then offered:

$$G_{ST} = D_{ST} / H_T$$

G_{ST} has become a standard metric used in quantifying genetic differentiation among populations. Values near zero are taken to indicate free exchange of individuals between the populations, whereas values near unity indicate isolation. It is one member of a family of such indices, all analogs of F_{ST}, which was derived by Wright in the 1940s. Where the original was suited to the biallelic loci of classic Mendelian genetics, the analogs are designed around other, more complex, situations (Meirmans & Hedrick 2011).

When originally proposed, G_{ST} worked well enough because the available laboratory techniques, such as isozyme electrophoresis, could only distinguish a few alleles per locus. However, the same non-linear relationship between the Gini-Simpson index and intuitive understanding of "diversity" that Jost (2006) emphasized for ecological diversity is seen in Nei's gene diversity. Advances in laboratory capabilities over recent decades have revealed very much greater genetic diversity than was formerly known, raising H_S into the range where H_T cannot be much higher, numerically, even if most of the species' genetic diversity is expressed among populations. Thus, since the mid-1990s it has been seen that G_{ST} (and most of the other F_{ST} analogs) can remain near zero even when there are no alleles in common among the various populations (Ryman & Leimar 2009; Heller & Siegismund 2009; Meirmans & Hedrick 2011).

---

[21] This use of H for heterozygosity should not be confused with the use of the same symbol elsewhere in this report for Brillouin's index of ecological diversity.

[22] This use of S should not be confused with the use of the same symbol elsewhere in this report for species richness.

Following his seminal work on ecological diversity, Jost (2008) addressed this obvious deficiency in G_{ST} by proposing the same ^qD indices, the Hill Numbers, that he had recommended for use with species-abundance data, though for geneticists he used the symbol Δ. For genetic diversity as much as for ecological diversity, the Hill Numbers vary as does intuitive understanding of diversity. Jost (2008) showed that calculation of relative differentiation among populations requires that each population be given equal weight and hence q should be set at 2, meaning that the measure of genetic differentiation should be $\Delta_{(q=2)}$ – the Inverse Simpson index of ecology. In the terminology of population genetics, $\Delta_{(q=2)}$ is the inverse of Nei's "gene identity" (also known as "homozygosity" and the complement of Nei's gene diversity). That was not exactly novel: Jost's (2008) $\Delta_{(q=2)}$ had been identified as the effective number of alleles as early as 1964. For applications where the sizes of populations matter, such as when examining the contributions of different hierarchical levels to total genetic diversity, the appropriate measure is $\Delta_{(q=1)}$, which is the same Exponential Shannon metric encountered above as ^1D (Jost 2008)[23].

As in the case of α, β and γ ecological diversities, Δ_T can be multiplicatively partitioned into Δ_S and Δ_{ST}. The latter is an "effective number of populations" and can take values from unity up to the number of real populations under consideration – equivalent to its analogue, β diversity, which is a count of "compositional units". To conform to the expectations and models of population geneticists, and as a replacement for G_{ST}, Δ_{ST} must however be transformed into an index with a range between zero and unity. Jost (2008) proposed the use of the inverse of Δ_{ST} (i.e. Δ_S / Δ_T), to which he gave the symbol D, and he did so despite seeing that his D has a lower bound equal to the inverse of the number of populations – much the same problem as one that later led him to attempt to overturn the concept of ecological evenness (cf. Jost 2010a – see Section 4.4 below). Jost (2008) also offered near-unbiased estimators of D for use with samples – an issue further discussed by Meirmans and Hedrick (2011). Since those estimators are algebraically complex, they are not repeated here but analysts faced with the estimation task should examine both papers.

Jost's (2008) proposals met at least as much swift resistance among population geneticists as his earlier ones (Jost 2006) did among ecologists (e.g. Ryman & Leimar 2009; Heller & Siegismund 2009). The resulting discussion has, however, transformed into further advances in, and understanding of, proposed metrics. Ryman and Leimar (2009) acknowledged the advantages of Jost's (2008) D as a measure of genetic diversity but claimed virtues for G_{ST} when the measure of differentiation is merely a step towards estimation of the migration rates between populations, a point that Jost (2009b) conceded. Subsequent authors have agreed that Δ or D is inappropriate when the aim is to draw demographic inferences but Δ remains the appropriate measure when quantifying allelic differentiation (Meirmans & Hedrick 2011; Whitlock 2011; Bird *et al.* 2011), for the same reasons that it (symbolized as ${}^{q}D$) is the most suitable index for quantifying ecological diversity, though the value of measuring allelic differentiation has been questioned (Whitlock 2011). Most have agreed that G_{ST}, and the other F_{ST} analogs, are the appropriate measures when estimating demographic parameters or migration rates, though Edelaar *et al.* (2011) have found them biased – a concern for some research work but not for biodiversity monitoring which need not use those indices.

---

[23] The roles within population-genetic theory of what ecologists would call the Gini-Simpson and Inverse Simpson indices have been well explored over decades, giving a particular status to $\Delta_{(q=2)}$, but Sherwin *et al.* (2006) have begun development of a parallel theoretical structure for the Shannon Index, which they therefore recommended as the common index of what they termed "biodiversity" at all levels from genes to landscapes.

Meirmans and Hedrick (2011), however, warned that Jost's (2008) D has to date only been developed as a measure of allelic variation at a single locus. Real-world determinations of genetic diversity almost always use data on multiple loci and further work was needed on multi-locus analogs of D. Subsequent efforts have considered the use of harmonic means of single-locus D values (Crawford 2010).

3.4 SUMMARY OF DIVERSITY THEORY

In summary, wherever there is a need to monitor temporal trends in, or map the spatial variations in, either ecological diversity or genetic diversity, the appropriate indices to use are those based on Hill Numbers, meaning Jost's (2006, 2008) ${}^{q}D$ and Δ – which are simply different symbols for the same metric, for application to different kinds of data. The series of Hill Numbers, and only that family of metrics, serves to quantify diversity in a way that is linearly related to how most users of the information perceive diversity to be.

For ecological diversity, ${}^{0}D$ (which is species richness, S) has utility, as does ${}^{1}D$, the Exponential Shannon index. The latter is preferred over ${}^{2}D$, the Inverse Simpson index, because ${}^{1}D$ is more responsive to rarities, which are of particular concern in biodiversity conservation. ${}^{1}D$ also has advantages in the familiarity of the Shannon index to most users. The optimal variant of Δ to use in monitoring and mapping of genetic diversity has yet to emerge. For diversity at a single locus, $\Delta_{(q=0)}$, meaning the count of alleles in the population at that locus, has obvious merit though (like ${}^{0}D = S$) it also has obvious limitations. $\Delta_{(q=2)}$ comes closest to being a familiar metric, being the inverse of the complement of heterozygosity, though $\Delta_{(q=1)}$ would appear to offer the same advantages as the Exponential Shannon index does for quantification of ecological diversity.

Where a need arises for measures of differentiation diversities, the multiplicative relationships explored by Jost (2006, 2008) and Tuomisto (2010a) are the appropriate ones, whether the aim is the β diversity of ecology or the Δ_{ST} of population genetics.

4 MEASURING ECOLOGICAL DIVERSITY

The three aspects of ecological diversity, richness, evenness and ecological diversity *sensu stricto*, may be generally uninformative and particularly so where the ecosystems of the open sea are concerned, as has been argued above (see Sections 2.3 & 2.4). They are little more than overly simplistic summaries of species abundances which discard much of the information in such data sets – information that can be captured and expressed by alternative approaches, some of which have been familiar to ecologists for three decades. Nevertheless, there is a broad expectation among non-specialists that ecological diversity forms the core of biodiversity and so, we suggest, index values representing that diversity must be included in national reporting of monitoring programs, whether presented as temporal trends, maps of spatial patterns or both (see Section 2.5). While those values must be presented, however, the presentation should be kept simple: there is nothing to be gained by elaborating almost-meaningless measures.

Ecological diversity *sensu stricto* was intended from the first as a single, scalar summary of a Species Abundance Distribution and could stand alone in that role. That has been justifiably criticised, however, since ecological diversity *sensu stricto* responds to both richness and evenness, frustrating its interpretation (see Section 2.1). At least two of the three aspects of ecological diversity *sensu lato* need to be determined if their variations are to be understood. It might be preferable to calculate and report richness and evenness, while setting aside their combination, which some commentators have suggested would aid clarity, simplicity and sensitivity (e.g. Dornelas *et al.* 2011; McGill 2011). In practice, however, evenness cannot be usefully determined without first quantifying both richness and ecological diversity *sensu stricto*, as is explained below (see Section 4.4). Yet, if values for the other two are available, determining evenness is trivially simple. Considering that the workload involved in calculating any of these index values is small, when compared to the effort required to gather the field data, we therefore recommend reporting all three of richness, evenness and ecological diversity *sensu stricto* whenever possible.

Contrary to frequent suggestions that measuring ecological diversity is straightforward, quantifying its indices in any meaningful way is a challenging task, especially for open-sea ecosystems, in which real variations (over time or space) are expected to be small and the potential for artifacts is large. In this section, the complications are explored and the indices to be used are discussed, leading to concrete recommendations for practical applications. The material is here presented in a discursive manner, with explanation and justification of the points made. A bulleted list of the steps to be followed in determining ecological diversity from DFO's existing sets of species-abundance data, while using our recommendations, is provided as Appendix II, which may be read alongside this Section 4. (Design of optimal, new survey protocols to gather data specifically for biodiversity estimation would involve additional, complex issues. That topic is not addressed in this report.)

For clarity of presentation, this Section is phrased throughout in terms of species richness and species diversity. The same concepts and indices can be applied to any other set of discrete units, including those at any level in ecological (e.g. habitats or landscapes) or taxonomic (e.g. genera or discrete intra-specific populations) hierarchies. Determining the diversity of genera or alleles should not pose any special difficulties beyond those considered here for species diversity, provided that knowledge of the system under study is sufficiently advanced for each record to be correctly assigned to the appropriate genus or allele[24]. Ecological diversity cannot, however, be well applied to those aspects of ecosystems that show clinal variations, rather than classification into discrete kinds. Clinally varying features may be arbitrarily divided into "types" (as is seen in many habitat classifications) but the resulting "type diversity" can owe more to artifacts of classification than to empiricism.

4.1 COMMUNITIES, TAXOCENES AND SURVEY PROTOCOLS

Theoretical debates over which diversity index to apply can obscure some important challenges facing their practical application – challenges which impose limitations on the interpretation of index values. Those need to be kept in the forefront whenever measurements of ecological diversity are considered.

4.1.1 Communities: Diversity of Where?

Concepts of ecological diversity first emerged in the era when community ecologists supposed that they could classify ecological units much as taxonomic ones can be. Diversities were thus to be measured for ecological communities, in the strict sense of that term. In practice, discrete, internally-homogeneous, recognizable ecological units are highly unusual in the biosphere and diversity must necessarily be quantified for much more arbitrary units. Thus, Magurran (1988) considered a "community" to be any group of interacting organisms within defined spatial boundaries – a definition consistent with those of other ecologists at that time and subsequently. McGill (2011) recently emphasized the importance of the interactions for the "communities" invoked in diversity studies, though he was reduced to demanding sufficient co-occurrence in time and space that there be "a reasonable chance of interaction" or just "an expectation that the members have the potential to interact with each other" – and that despite his admission that, in some studies, "communities" could span continents or millions of years. Indeed, he went on to question the existence of "communities" in any real sense but nevertheless saw value in considering the diversity "in local regions of space-time" (Maurer and McGill 2011).

---

[24] It has been suggested that ecological diversities at higher taxonomic levels should be used in place of species diversity (thus employing the higher taxa as "taxonomic surrogates" for species), greatly reducing the costs of processing samples (ICES 2010; Bevilacqua *et al.* 2012b). That is unlikely to be advantageous when working with Canadian aquatic vertebrates, since identifications are straightforward while few genera contain many species, but there might be substantial savings with some invertebrate, phytoplankton and microbial taxocenes. The approach has proven advantageous in studies of benthic polychaetes in northern European seas (Olsgard *et al.* 2003) but Bevilacqua *et al.* (2012b) have recently examined "taxonomic sufficiency", the ability of studies of higher taxa to reveal no less of anthropogenic impacts than species diversity does, and have found that performance is poor when there is an average of more than 2.5 species per higher taxon (suggesting that surrogates are sufficient when most higher taxa are represented in the community under study by a single species, which negates most of the cost saving). Further exploration of "sufficiency" is needed before opting for monitoring or mapping of "taxonomic surrogates".

In practice, the interactions amongst the organisms in some chosen unit cannot usually be known without very extensive research, meaning that they are generally ignored and that ecological diversity is most often quantified simply with reference to the biota found within some area of interest – which area is often defined by rather arbitrary boundaries. Maurer and McGill (2011) termed the organisms within such a unit an "ecological sample" (distinct from the "empirical sample" gathered by a scientist). That suggestion risks terminological confusion, and so will not be adopted here, but it is a useful reminder that the "communities" for which diversity may be estimated are but sub-units of larger ecological systems. That is not necessarily a weakness but it does throw an extra responsibility onto those conducting monitoring programs and interpreting the resulting data: since the units for which diversity is measured are at least partially arbitrary, rather than biologically based, there is a risk of artifacts being introduced by the selection of community boundaries. Analysts must be aware of that risk and must ensure that any reported conclusions arise from empirical data and not from arbitrary decisions embedded in the design of the monitoring program.

To give but one example: Should it be necessary to compare fish diversity on the flat top of an offshore bank with that in an inshore area of small banks, basins, channels and ridges, it might be appropriate to compare the communities within 10-mile squares, when the inshore area would likely have many more species because of the greater variety of habitats. Alternatively, the comparison could be of only the communities on the bank-tops, when the larger offshore banks might hold more species simply because they are larger. If the issue at hand was selection of the smallest MPA that could encompass representatives of the most species, then the former design might be appropriate, whereas if the concern was over the structure and function of bank ecosystems, the latter one would seem more suitable. In either case, defining the communities to be considered by inappropriate boundaries could lead to entirely erroneous conclusions.

In trend monitoring, it will always be desirable to maintain the spatial bounds of the chosen community with minimal change over time, to avoid adding noise to the data and so obscuring any temporal signal which may be present. The data can nevertheless be regarded as being gathered from multiple "communities", divided from one another in time, although overlapping in space. For greater clarity, such units are here termed "temporal communities", though for the purposes of diversity measurement they are no different to other communities, including those which overlap in time but are discrete in space.

4.1.2 Taxocenes: Diversity of What?

There is a superficial attraction to the notion of measuring the diversity of the entire biota of a chosen community but ecological diversity in that comprehensive sense has rarely, if ever, been quantified. Should the attempt be made, the relative abundances of metazoans and multi-cellular plants would be swamped by those of the microbiota and the resulting index values would be little different from what would have been achieved by a study confined to bacterial and viral diversity alone (cf. Hurlbert 1971; ICES 2010). Even with the microbiota excluded, most measures of "overall" diversity in aquatic environments (if developed from data on numerical abundances) would be dominated by microalgae in the water column and meiofauna in the sediments (cf. ICES 2010), to the effective exclusion of the megafauna that, especially in the open sea, are more likely to respond to anthopogenic disturbance. Hence, even as an ideal (though more so in practice), ecological diversity can only usefully be quantified for a defined group of species. Conventionally, the group should be a "taxocene", meaning all of the members of one taxon that are present within the area of interest (Pielou 1974, 1975, 1977). In the case of the avifauna, members of the Class Aves, that works well. The vernacular "finfish" does not, however, refer to a particular taxon nor even to a polyphyletic grouping but rather to an amalgamation of aquatic vertebrates that respire through gills in their adult stages, including the totality of some taxa added to parts of others. Thus, it may be necessary to evaluate not the ecological diversity of a single taxocene (e.g. all members of the Teleostei) within the chosen community but that of a set of taxocenes, which might include a combination of teleosts, sharks and lampreys, while excluding the tetrapod vertebrates. Similar complications can arise with some groupings of aquatic invertebrates.

In practice, the "taxocenes" used in measurement of ecological diversity are commonly confined to certain life stages of the organisms of interest. "Angiosperms" typically exclude ungerminated seeds (which are sometimes separately considered in the diversity of "seed banks"), "birds" generally exclude eggs and nestlings, while "insects" may be confined to the flying adults. In marine applications, the meroplanktonic eggs and larvae of species that are benthic or nektonic as adults are likely to be excluded from diversity measurements, unless there is a particular focus on, for example, ichthyoplankton diversity. All such restrictions on the chosen taxocene should be explicitly stated, though too often they are not. Sometimes, attention is confined to particular life stages in order to add ecological meaning to the measurements. Avian diversity, for example, may be restricted to breeding birds, thus eliminating the many transients seen with such a mobile taxon (Magurran 2011). That restriction would, however, lead to diversity values that ignore sometimes-important winter feeding aggregations and the use of critical "way station" feeding grounds along migration flyways. Similar weaknesses in marine applications led ICES (2011b) to caution against such restrictions on the scope of a taxocene. A more problematic option is to restrict the taxocene to individuals that fall within some chosen size range. In studies of benthic infauna sampled by grabs, for example, it is common to sieve samples and examine only those animals retained by a sieve of some chosen dimension (Gray 2000). While valid comparisons can be made, across time or space, using such size-restricted data (provided that the size-selecting methods are standardized), those comparisons may not be ecologically meaningful: an area with markedly higher species diversity amongst benthic individuals larger than 1 mm might show markedly lower diversity if the samples had been sorted over a 0.5 mm sieve.

For special purposes, it is sometimes appropriate, or even necessary, to ignore taxonomy entirely and measure the diversity of groups other than taxocenes. Pielou (1975) noted the anomaly that aquatic ecologists often determine the ecological diversity of all fish in a lake, including predators and prey together, whereas woodland ecologists would not consider combining insectivorous birds and their prey into a single diversity measure. Hurlbert (1971) suggested splitting those taxocenes which span more than one trophic level, in order to increase the ecological homogeneity of the unit of study. While challenging in many aquatic communities, since many species ascend through multiple trophic levels within the lifetime of a single individual, limiting the defined "taxocene" for a particular determination of diversity to a single trophic level is certainly worth considering. Hubbell's (2001) "Unified Neutral Theory", for example, addresses the diversity of what he termed "communities" but were closer to trophic guilds. In more typical monitoring and mapping work, it is sometimes possible to relax the linkage to a taxocene and to measure the diversity of some partially-arbitrary grouping of organisms, such as "groundfish". There is, however, a danger that artifacts will be introduced, since any grouping not firmly anchored in taxonomy requires arbitrary choices of what species to include or exclude, when the inclusion or exclusion of rarities could have major effects on the estimated values of the diversity index. Thus, while McGill (2011) did not limit the use of diversity measures to literal taxocenes, he did warn that the list of species considered must be, in some meaningful way, exhaustive, with nothing arbitrarily omitted. It would be easy to "prove", for example, that species richness is invariant, simply by excluding from the calculations all of those rarities which may be absent from some waters or which may disappear over time. Thus, care is needed when defining a chosen "pseudo-taxocene" and when drawing conclusions from the data on that unit. It is essential that the organisms included in each calculated value of ecological diversity, and hence those excluded, are clearly defined and that the same grouping of biota is used in any subsequent measurement to which a first one may be compared: conclusions based on a comparison of teleost diversity at one time with the diversity of all finfish at some later time would be misleading – unless all of the finfish present are teleosts.

The taxocene must be selected such that sufficient taxonomic knowledge and technical resources are available for specimens to be reliably and consistently sorted into their species. That is little constraint on measurements of the diversity of Canadian vertebrates but many marine invertebrates are more challenging (a significant problem for studies of benthic diversity in European seas: ICES 2010). Where existing databases are used, rather than new fieldwork being planned in support of diversity measurement, this becomes a requirement that all specimens in the chosen taxocene were reliably and consistently sorted when the data were collected. The lack of past consistency has been identified as at least a potential problem with both the Scotian Shelf groundfish trawl surveys (Shackell & Frank 2003) and the equivalent Scottish surveys in the North Sea (Greenstreet 2008), since the level of taxonomic skill carried aboard the survey trawlers, the extent to which large catches were fully sorted and / or the level of interest in devoting time to identifying every specimen have all varied over decades. Much the same can be expected of other survey series. While they are problems with data on finfish, temporally varying degrees of taxonomic expertise and effort are even bigger challenges with datasets on the invertebrates taken on the same surveys (ICES 2010). Fortunately, a strict focus on a particular taxocene usually aids the onerous process of specimen identification, laboratory staff typically being most familiar with the members of one major taxon. Judicious selection of which taxocene to monitor can also provide measures of diversity that are more relevant to immediate management concerns.

Fishery-survey databases often contain further problems for the diversity analyst, such as "difficult" taxa being sometimes recorded by their generic names and sometimes by those of higher taxa, which can necessitate merging records of the affected lower taxa into the higher ones (ICES 2010). Familiarity with the idiosyncrasies of a particular dataset is always essential before it can be reliably used in the measurement of diversity. In many cases, that familiarity will show a need for extensive data editing, even if the dataset has already been brought to the standard required for the quite-different purposes for which it was originally collected.

When mapping ecological diversity, it should never simply be assumed that areas with a greater diversity of some chosen taxocene are necessarily more diverse for others. While a greater variety of habitats might be expected to harbour enhanced diversities of many taxocenes, or that higher prey diversity should support more varied mixtures of predators, in practice the areas showing higher ecological diversity are not necessarily congruent across taxocenes (Magurran 1988). Congruence is often, though not always, seen in large-scale studies (e.g. continent-wide) but it tends to be weak at regional scales (e.g. Heino *et al.* 2005, 2009; Heino 2010). In a recent marine example, illustrative of the complexities that may be encountered, Ellis *et al.* (2011) examined an area of sandbanks and intervening channels off the English coast, finding that the channels held more diverse communities of fish, epibenthos and macro-infauna than did the bank crests, but that the meiofaunal nematodes showed the reverse trend. Olsgard *et al.* (2003), in contrast, found that polychaete sample richness was highly correlated with overall invertebrate macrobenthos richness, perhaps because the varied feeding modes of polychaete species make that taxon responsive to the multiple factors that shape the richness of other phyla. The richness of the Order Terebellida proved to be a good indicator of both overall polychaete richness and that of the macrobenthos as a whole (Olsgard *et al.* 2003). Thus, where the available data permit, values of ecological diversity should initially be determined for multiple taxocenes but, in some cases, congruence may allow later economies – after that congruence has been demonstrated through case-specific trials.

There has been some discussion of means to combine the diversity values for multiple taxocenes (e.g. ICES 2011b). That is almost certainly a mistake, since any combination would hide correlations among taxocenes that share patterns of variation in time or space, while masking change or spatial variability where taxocenes show contrasting patterns. Unless diversity is measured for each of so many different taxocenes that the results become overwhelming, it would be better to consider and present them separately.

Hurlbert (1971) suggested that, when selecting a taxocene, the chosen taxon should be one low in the Linnean hierarchy (e.g. a family or order) so that its members would respond similarly over time or space and thus contrasting trends would not be masked. He recognized, however, that narrowing too far (e.g. to genus or even family) would make species richness within the taxocene too low for any changes to be detected. There may be ecosystems in which autecology, and hence responses to disturbances, closely mirror taxonomy but that is rarely evident in relatively-depauperate Canadian communities, removing much of the point of that proposal. We suggest instead the examination of functional traits diversity whenever ecosystem function is of primary interest, confining ecological diversity to its role as a summary of species-abundance data.

Some authors have been concerned about an entirely different issue: the limited meaning of a comparison among values of ecological diversity when the communities concerned have few species in common within the chosen taxocene (e.g. Gotelli & Colwell 2001, 2011). It is true that such a comparison can be all but meaningless, save when answering certain research questions. That, however, is a basic defect of the concept of ecological diversity, with its blindness to species replacement and its focus on reducing species-abundance data to a scalar value, or at most three of them. Perhaps fortunately, the problem will rarely arise in practical applications in marine ecosystems since it is rare (if not unknown) for any one survey protocol to be used across such a wide variety of communities that some have few species in common with others – while quantitative comparisons should not be made between the diversities of communities surveyed using different protocols (see Section 4.1.4). If examples are encountered of communities with comparable data and yet limited overlap in species compositions within the chosen taxocene, then emphasis should be given to analyses which respect species identities, unless the focus is on research questions concerning ecological diversity itself (as distinct from ones that use diversity to study something else).

Within a chosen taxocene, and supposing the issue of interest is species diversity, it is usual to bin the data by species but there is no special reason why sub-species, or even local populations, should not be distinguished, provided that artifacts are not generated thereby. Pielou (1975) noted a related problem with those taxa that form "hybrid swarms", with only vague and uncertain distinctions between one species and the next. (While that is primarily an issue in terrestrial plant systems, the cichlid fishes of the African Great Lakes pose a similar challenge, albeit not for Canadian ecologists.) Where there is a desire to link diversity measurements to ecological theory, there is a further complication with species that can exploit more than one niche and thus act, ecologically, as multiple species (Pielou 1975). That is a challenge with the many teleosts that ascend the trophic ladder as they grow.

4.1.3 Temporal Cycles: Diversity of When?

A further problem long known to quantitative ecologists is that diversity varies over a range of temporal cycles, quite apart from the inter-annual trends that are of interest when tracking changes in ecosystems (e.g. Pielou 1974, 1975). Hence, just as any estimate of ecological diversity can only be made for a specified area, so it can also only be made for a specified time – either a single block of time or a period that recurs cyclically. Terrestrial plant communities containing annual species necessarily see seasonal cycles in ecological diversity and the same can be expected in both planktonic and benthic communities, whether of plants or animals. Indeed, much of the plankton, like many terrestrial insects, have only short seasons of abundance or activity within each year. Animal migrations also drive seasonal cycles in ecological diversity within spatial units smaller than the migration circuit, even for species with life expectancies of multiple years. That pattern is familiar with the avifauna of terrestrial ecosystems but long-distance movements are much more prevalent amongst marine animals, making seasonal variations in local diversity the norm, rather than the exception (cf. ICES 2010). Depending on time, place and the behaviours of various species, ecological diversity can show pronounced changes over periods as short as a few days.

Cycles with even shorter periods can be equally important. Many flying insects are active, and hence vulnerable to light traps, only during one part of each night, with some species being taken at dusk, others at dawn and yet others around midnight (Pielou 1974). It follows that measured diversity can vary over a diel cycle. The same may emerge in the sea if, for example, diel vertical migrations carry some species above or below the depth-range sampled by a survey or bring pelagics down to where they are vulnerable to bottom-tending gears. Inter-tidal and very shallow sub-littoral areas see similar changes on still shorter periods, primarily semi-diurnal. It is sometimes possible to deploy survey gears around the clock or else to average across samples taken at different times of day (though "communities" thus studied may contain species that never interact with one another, being kept apart by diel behaviours: cf. Pielou 1975). The general principle remains: calculated values of diversity indices for a chosen community are specific not only to the selected taxocene but also to the particular season and, sometimes, phase of the diel and / or tidal cycles when the data were gathered.

Measured ecological diversity may also vary across time non-cyclically, especially when the capture or recording process depends on the behaviour of animals. Weather effects on the recording of insects with light traps are a textbook example (Magurran 1988) but the same might be expected with observations of seabirds or marine mammals from shipboard and to some extent with the performance of nets when trawlers encounter heavy seas. Low water temperatures are known to affect lobster foraging and hence their susceptibility to trapping, so differential temperature responses amongst benthic crustaceans could affect species' relative abundances recorded in a trap survey. Some benthic infaunal species move vertically within the sediment, in response to various drivers, potentially moving into and out of the range of the chosen sampling gear. At a temporal scale between weather patterns and climate change, such patterns as the *El Niño / La Niña* exchange and the similar but less-known shifts in the North Atlantic linked to the North Atlantic Oscillation certainly affect marine communities and presumably their ecological diversities.

In other settings, survey periods may be too short to capture key ecological processes. Forest ecologists face challenges with their data collection being very brief when compared to the life of a tree (Pielou 1975), let alone the process of succession in a patch of woodland. Similar complications might arise in studies of corals or some macroalgae.

Overall, it is important for the analyst to be aware of the effects of time on diversity measures and to ensure that survey timing yields data that are relevant to the questions to be addressed. Care is needed both in the design of field surveys to avoid artifacts driven by short-term temporal changes and in the interpretation of the data to ensure that conclusions are not based upon those changes – unless the topic of interest concerns diversity change over short durations.

4.1.4 Survey Protocols: Diversity How?

In some of the settings which originally interested quantitative ecologists, it is possible to count every individual organism of the chosen taxocene within each survey area. The angiosperms within a 1 m^2 quadrat, for example, can be fully counted (setting aside complications arising from growth forms that confuse the delineation of individual plants) and identified to their species, at least if ungerminated seeds are excluded. Something similar may be possible in aquatic environments if visual survey methods (by SCUBA diver or ROV) are combined with a taxocene definition that makes a census of either quadrat or transect viable. Nevertheless, obtaining data on the species present and their relative abundances is usually dependent on some survey technique that is necessarily selective. Even trained ornithologists observing through binoculars, who may be able to record all bird species present, cannot determine their true relative abundances since some types are more cryptic than others. Marine biologists must normally sample with gears that are highly, sometimes very highly, species- and size-selective – the latter affecting recorded relative abundances of species as their size compositions vary across time and space. The near-inevitability of selective sampling is an ever-present fact that the sub-discipline has struggled with from its beginning. Indeed, most aquatic ecologists are likely more familiar with the constraints of survey selectivity than many specialists in ecological diversity appear to be.

While it is usual to consider gear selectivity, the challenge is actually broader and more akin to what is sometimes known in fisheries circles as "partial recruitment". For each species in the taxocene and each size-class in that species, the selection at issue relates to the ratio of a relative abundance in the survey dataset to a relative abundance in the community (as that has been bounded for a particular analysis). That ratio is partly a matter of how the chosen survey gear selects among the mix of species and sizes where it is deployed and partly one of how and where the gear was deployed across the community, as well as any additional selection of specimens that may have occurred during processing of catches or editing of data (e.g. pooling or discarding of species of no resource value). Bottom-trawl surveys, for example, usually avoid rough seabeds and those with steep slopes, to minimize gear loss. Benthic grabs will not "bite" when dropped onto very hard seabeds and grab-drops are typically only accepted as valid (thus contributing to datasets) if they are made where sediments are soft enough for a bottom sample to be taken. Thus, it is not merely the selection processes of the survey gear that are of concern but rather those of the entire survey protocol.

With enormous effort, it is theoretically possible to inter-calibrate amongst different protocols and gears, so that data from them can be pooled. In some cases, it may even be possible to estimate selectivities, and thus to determine absolute abundances, but even the inter-calibration can only be done for species commonly encountered – and hence for which sufficient data can be gathered. Experience with the Department's groundfish resource surveys, when old ships were retired and new ones towing nets of different designs were introduced (e.g. Fanning 1985; Benoît & Swain 2003; Fowler & Showell 2009), suggests that even such limited inter-calibration of the data on the major commercial species is problematic. Ecological diversity is, however, sensitive to the presence of rarities and, for species rarely caught, it is grossly improbable that inter-calibrations could be determined with useful precision.

Buckland *et al.* (2011b) have recently addressed this issue, which they labelled as one of "detectability", from within the perspective of specialists in ecological diversity. Unfortunately, they postulated a dichotomy between determining the diversity of a chosen taxocene within a community and determining that of the taxocene after filtration through the selectivities of the adopted survey protocol, with that filtered diversity assumed to represent the diversity of the community. Correctly rejecting the latter, untenable assumption, they were led into a doomed attempt to estimate the selectivities of the various species observed by a visual transect survey for terrestrial birds (Buckland *et al.* 2011b). In practice, they failed to fully address the one aspect of "detectability" that they attempted to measure, did not touch on a second that they noted and entirely missed a third[25] – which illustrates the scope of the challenges in selectivity estimation. In the end, Buckland *et al.* (2011b) conceded that it may be acceptable to ignore the effects of "detectability", so long as the biases are consistent throughout. We would argue that, with rare exceptions, there is no other alternative. The mistake is to suppose, as Buckland *et al.* (2011b) considered doing, that the species' observed relative abundances, filtered through some survey protocol as they must be, represent those of the same species in the community. In general, they do not and cannot. Rather, in almost every case, a measured ecological diversity is not the diversity of some chosen taxocene within the defined spatial and temporal bounds of a chosen community but rather the diversity of that taxocene as filtered through the selective properties of the adopted survey protocol – what ICES (2005) described as "diversity as observed in the image of the community as viewed through the sampling gear". Such a diversity of a "filtered" taxocene is (usually) all that is available and hence must be the unit of interest. It follows that almost all measured values of ecological diversity are specific not only to the chosen community and taxocene but also to the particular survey protocols used in gathering the data. There is nothing new in that conclusion: some of the classic data sets of quantitative ecology concerned insect communities sampled by light traps, which only take those species that are attracted to the light.

In trend monitoring, where the interest is on temporal change or its absence, the dependence of measured diversity on survey protocols may be no limitation at all, provided that the protocols can be maintained constant. That is a familiar requirement within the Department but an onerous one nevertheless: there is always a tendency for standardization to drift over time unless rigorously enforced (cf. McCallum & Walsh 1997, 2001). Even with the greatest dedication, major changes in protocols cannot always be avoided, such as when groundfish-survey trawlers have to be replaced with new ships. Such pronounced disruptions may call for a time series to be split, without any attempt to draw comparisons between the pre- and post-change eras. Most certainly, usefully accurate trends in ecological diversity can only be determined across periods for which data with consistent biases, usually meaning data from a single survey series, are available[26]. Moreover, few surveys to date have been either designed or operated for diversity monitoring. Rather, monitoring of ecological diversity usually draws data from surveys intended for other purposes, particularly monitoring of resource biomass. Not infrequently, such surveys require minor protocol modifications, deemed not to affect the primary objectives of the fieldwork, for practical reasons. A change in the material used in net twine or modification of the layout of a trawl's groundgear, for example, might not much affect estimation of the biomasses of the principal resource species, while the consequences of the modification for scattered catches of rarities are too easily ignored. Over time, a trawl survey may gradually come to avoid areas of harder bottom, as the survey team accumulates records of "hangs" and thereafter rejects any randomly selected station positions that pose risks to the nets, potentially creating a marked downward trend in the survey's selectivity for hard-bottom species. Of present importance, such mild modifications in methodology frequently pass without published record and are soon forgotten by those running survey programs (if only through staff turn-over breaking the chain of institutional memory) – meaning that key details too easily remain unknown to analysts calculating the values of diversity indices.

---

[25] As to the third aspect: Birds flying over the transect during the observations were recorded but Buckland *et al.* (2011b) set those data aside – as they must be when calculating species densities from such a survey. Hence, those species which respond to observers by flying or which simply spend more time on the wing were under-represented in the analyzed data, though likely over-represented in the full dataset.

Other unanticipated challenges can arise when data from resource-biomass surveys are used as a foundation for diversity measurement. ICES (2010) noted, for example, that many epibenthic species can become hung up in the meshes of survey trawls, to be recorded in the catches of later sets when they finally wash into the codend. The same happens with teleosts, albeit to a lesser extent. Since the amounts of biomass that are thus misrecorded are small, while the biomass estimates are computed as regional totals, survey protocols tend to be rather casual in their processes for minimizing the resulting data corruption. Values of the species richness of single samples can, however, be materially affected by even one individual of a species being recorded as taken in the wrong set. In some applications of diversity mapping, such errors can lead to incorrect conclusions.

---

[26] There have been attempts to determine trends in diversity over longer periods. Palaeontologists have even studied changes in ecological diversity through geological time (Rosenzweig 1995), despite the considerable variations in fossil survival rates – a form of selectivity in a survey "protocol" that relies on fossilization as part of its process for obtaining data. More recently, there have been attempts to reconstruct ecological change over centuries by assembling whatever assorted data may be available. Used carefully, such approaches can reveal major changes in diversity but minor ones will be lost in the uncertainty surrounding the relative values of the selectivities inherent to different types of data. When engaging in such work, analysts bear a responsibility to ensure that consequences of temporal change in selectivities are not misrepresented as changes in diversity, and equally to avoid real trends in diversity being masked by those in selectivities.

It is also worth noting that data from non-standard survey units (e.g. trawl sets that are prolonged beyond the standard duration) cannot readily be standardized, after the event, since the selectivities of some survey gears change with the duration of their deployments. Longer tows with trawl nets can result in full codends, which is also an effect of richer fishing, the blocked meshes tending to retain small fish that would otherwise have escaped. Large fish capable of out-swimming a trawl can be caught if the set is prolonged beyond the point of exhausting the fish. Conversely, baited traps and longlines become less efficient over time as the odours of the bait are lost. Gillnets may become more visible to fish as they become fouled. It would be grossly impractical to estimate, for every species of interest to a diversity analyst, the magnitude of all of the possible changes in selectivities driven by such deviations from standard protocols and hence it is important to maintain the units of survey effort as constant as is practical, while rejecting any that fall far from the intended standard – perhaps exercising more rigorous rejection than is required in biomass estimation. Diversity analysts should further be aware that available survey databases may have already been subjected to adjustments intended to standardize the inevitable non-standard units, potentially in ways that (while suited to the primary objectives of the survey) could create artifacts in diversity measurements.

Thus, great care will always be needed when drawing on existing survey data for examination of long-term trends in ecological diversity. The diversity analyst bears a responsibility to closely investigate any dataset proposed for analysis, in light of the intended analyses and potential conclusions, to ensure that the inevitable changes, across time and space, in survey protocols (which may be in part unwritten "protocols" established by regular practice) have not resulted in sufficient alteration of selectivities as to invalidate any conclusions that may be drawn. Unlike the situation in terrestrial and some freshwater or coastal ecosystems, changes in ecological diversity in the open sea are expected to be minor. If real, they could easily be masked by artifacts of altered selectivity. If not present in the real ocean, their appearance in the data could easily be created by such artifacts.

The ready acceptance that measured diversities are usually those of taxocenes as filtered through the selectivities of survey protocols may be less easy in applications other than trend monitoring. Ecological research that compares empirical data to theoretical expectations could be severely compromised. Of more immediate concern in this report: when mapping spatial patterns of diversity, it is clearly important to use the same survey protocol throughout the area covered by any one map. It may be acceptable to conclude that coral reefs have higher teleost diversity than do cold-temperate offshore banks, despite the former being surveyed by SCUBA divers using visual techniques while the latter require trawl surveys, but if so it is only because the difference in species richnesses is so great as to overwhelm the artifacts generated by the different means of collecting data. Conclusions about the finer differences between, say, the bivalve molluscan *faunae* of two offshore banks cannot be built on sampling with different grabs, the one bank with a small grab that only takes near-surface species but the other with larger gear that also takes the deeper-burrowing infauna. More subtle but equally serious complications may arise, such as when some parts of a variable seabed have extensive patches of hard bottom, bearing a rich epifauna that is inaccessible to the chosen grab, while the surrounding soft sediments tend towards spatial uniformity in macrobenthic ecological diversity. The map developed from such a survey might suggest a quite misleading spatial homogeneity, dominated by the same infaunal community throughout.

These are not avoidable challenges. Rather, it is important that both those who measure ecological diversities and those who use the results remember that the numbers generated relate to taxocenes within communities not directly but only through the filter of survey-protocol selectivities. It is no less important to ensure that any conclusions drawn are not founded on the inevitable artifacts that those filters create. Unfortunately, those points have too often been forgotten. Magurran (2004) for example, faced with the near-inevitability of selective data collection, recommended the use of a wide array of sampling techniques in order to detect all species present by ensuring that "all potential niches are searched". That is sensible advice for those seeking to prepare a comprehensive species list but muddling the data from multiple sampling protocols would prevent either the preparation of either a Taxon Sampling Curve (see Section 4.2.1), which would block estimation of species richness in most situations, or meaningful species' relative abundances, eliminating the use of metrics of ecological diversity *sensu stricto*. Similarly, Callaway *et al.* (2007) sought to examine long-term change in the epibenthos of the North Sea, which necessitated use of data collected by assorted gears as far back as the Edwardian era. They correctly understood that such data could not be analyzed using any of the conventional indices of ecological diversity but erred in supposing that the sole problem was the impossibility of reducing the various samples to a common standard of sampling effort. Callaway *et al.* (2007) thus opted to use Clarke and Warwick's (1998) "taxonomic distinctiveness" metric, which is independent of sample size (see Section 5.3). While that solved one problem, it did not address the consequences of the different selectivities of the sampling gears, which undermined all of the conclusions of the study.

It should perhaps be stressed that the selectivities of the chosen survey protocol for the various species in the taxocene of interest are not merely unknown and variable from protocol to protocol but also different from species to species. One important consequence is that a decline in abundance to zero survey records cannot be interpreted as synonymous with a decline to some sort of "functional extirpation". Species with low "detectability" may disappear from the records of a particular survey while still abundant in the surveyed ecosystem, while other species (having much higher "detectability") may continue to be routinely recorded despite severe depletion. The abundance densities at which particular species disappear from the records of particular surveys have no ecological nor conservation significance. Hence, failure to record a particular species amongst the catches of a particular survey, even if that failure persists over multiple years, is not evidence of a decline in species richness (though it could represent a fall in evenness). As explained below (see Section 4.2), richness must be determined in terms of the number of species that a survey could record, not the number that it has recorded.

The constrictions of ecological diversity measurement onto particular taxocenes and particular survey protocols interact: the taxocenes for which diversity can be measured are necessarily those for which appropriate sampling gear can be deployed – often those for which appropriate gear has been routinely deployed in the past and for quite other purposes. The nuances of the taxocene definition can be particularly gear-dependent. The “fish” taken by a groundfish trawl survey, for example, exclude all planktonic stages and very likely exclude all large pelagic species also. There is sometimes a desire to exclude from the data used in diversity measurement those species that are only incidentally caught by the standard gear. Should a tuna or swordfish ever chance to be taken by a bottom-trawl survey, there may be good reason to discard the aberrant specimen from the calculations. That can go further, however, to the extent of rejecting all species deemed to be pelagics from what is conceived to be a dataset on demersal fish. If that is done, great care must be exercised to ensure that subsequent comparisons of diversity values do not hinge on artifacts created by the selection and that the restricted “taxocene” is still relevant to the issues at hand.

Indeed, the issue of relevance of filtered taxocenes must be considered even if no artificial restrictions are applied. Fishery resource surveys are usually conducted at fixed times of year, to minimize the inflation of uncertainty through including seasonal changes in catchabilities. In consequence, they provide “snapshots” of the distributions of migratory species which need not be representative of where those same fish spawn, feed or are fed upon. Moreover, the species that are regularly taken in typical trawl surveys usually exclude those that are at most risk of extirpation or extinction – the coastal species not found on offshore trawling grounds, the anadromous species which are pelagic when in the sea, and the large top predators that readily out-swim trawl nets (cf. ICES 2010, 2011b). A combination of preliminary analyses and careful consideration is needed to ensure that whatever spatial or temporal signal may be found in diversities determined from a particular dataset is relevant to the management issues at hand. Those are even more necessary to determine the relevance of a null result: an absence of any detected signal. The sorts of biodiversity loss that are of policy concern could pass entirely undetected by a groundfish trawl survey, for example (cf. ICES 2011b). More generally, it should never be assumed that analysis of existing survey data series will suffice to measure those facets of biodiversity that are of policy concern.

It is not the role of this report to critique particular past attempts at measuring biodiversity. However, our conclusion that almost all such measurements concern taxocenes only as they are filtered through survey protocols stands counter to much that has been assumed or implied in the literature on diversity – even though it is not unique to us (e.g. ICES 2005). Some of the inferences that we have drawn, especially that the measurements are only inter-comparable across space or time within the confines of a single survey dataset, depend not on the existence of selectivity filters but on their differing substantially among survey protocols. Others might argue that the differences are small enough, relative to spatial or temporal variations in ecological diversity, to permit meaningful conclusions to be drawn from comparisons across multiple survey series. While those who seek to make such comparisons should establish their validity on a case-by-base basis, it may be worth considering here some attempts to derive measurements of ecological diversity from groundfish-survey data – a source that is among the best available for temperate continental-shelf ecosystems, including in Canadian waters.

Fraser *et al.* (2007) estimated the selectivities (which they termed "catchability coefficients") of North Sea trawl surveys for multiple size classes of each of 29 demersal species, as a step towards estimation of overall demersal fish biomass. Five of the 29 (including cod and plaice, *Pleuronectes platessa*) were routinely assessed and their selectivities were found by comparing survey abundance estimates to stock-assessment estimates (by species and size class). For 21 other species, selectivity was estimated as a weighted average of the weighted selectivities for plaice and cod of the same length class. In each case, the weights were the ratio of the per-set catch (of the species and size-class in question) in the survey trawls to those in beam trawls – the ratios being divided by equivalent ratios for plaice or cod, as appropriate (Fraser *et al.* 2007). (The three remaining species lacked adequate data for that approach and had selectivities estimated by even cruder methods.) Selectivities thus estimated may be fully adequate when developing estimates of overall community biomass, since that is dominated by the few species for which stock-assessment results are available, while errors in other estimates will tend to average out. Fraser *et al.* (2008), however, used those same selectivities in a study of ecological diversity. If that were a valid approach, it would go some way towards disproving our claim that inter-calibration across survey series is not viable – though the restriction to just 29 species greatly reduced the challenge of selectivity estimation, at the price of eliminating meaningful estimation of species richness. Moreover, measures of ecological diversity *sensu stricto* are sensitive to the relative abundances of minor species. Thus, Fraser *et al.* (2008) required reasonably precise estimates of the selectivity of each species, rather than the adequate average of those estimates which was all Fraser *et al.* (2007) needed. Those authors' second paper offered no support for the proposition that simple weighted averages of plaice and cod values, without regard to whether the species of interest is a flatfish or a roundfish, can meet that required standard and hence negate our conclusion that such estimation is impractical at best.

Fraser *et al.* (2008) went on to use the metrics ^0D, ^1D and ^2D, though they made no attempt either to determine the asymptotic values nor to rarefy (see Sections 4.2 and 4.3 below), relying instead on Greenstreet and Piet's (2008) conclusion that pooling the data from 20 sets of the International Bottom Trawl Survey is sufficient to determine the relative rankings of the asymptotic values of the diversity indices, when the taxocene in question is reduced to just the principal demersal species. Within that limitation, they reported very profound differences between the data derived from beam trawling, which showed high diversity along the British coast, and that from otter trawling, which found a band of elevated diversity running east from Scotland to Denmark and another such patch in the Southern Bight. The attempted corrections for selectivity generated a third dataset that gave yet another picture, with elevated diversity across the southern North Sea, while the waters between Scotland and Norway appeared relatively depauperate. Hence, even for one limited group of species, surveyed with two kinds of trawls or one trawl with adjustments for selectivity, maps of measured "diversity" can be so radically different that no management decisions could usefully be built on them.

For the Scotian Shelf and Gulf of Maine and for a "taxocene" of twenty selected large-bodied groundfish, Shackell *et al.* (2012) prepared a time series of evenness measurements (using what they described as "Pielou's Evenness Index" – probably meaning J' which, in their application with fixed S, is directly proportional to Shannon's H'). In doing so, they disregarded the change in the Canadian surveys from *A.T. Cameron* towing a Yankee #41.5 trawl to *Alfred Needler* working a Western IIA, and also the difference between the two Canadian trawlers and their nets, used in the eastern portion of the study region, and the U.S. vessel *Albatross* and her Yankee #36 net used in the west (albeit with broad overlap in areas surveyed). Shackell *et al.* (2012) justified that combination of ostensibly non-comparable datasets with a tri-partite argument: an appeal to precedents set by earlier authors, the lack of data for inter-calibration, and a broad regional coherence in single-species temporal trends (across the same three survey series) that Nye *et al.* (2010) had observed in fifteen of the twenty species. None of those three can give any support to the implicit assumption that the artifacts in evenness values which result from ship- and net-effects were too minor to affect Shackell *et al.*'s (2012) conclusions. However, inspection of their results shows no pronounced step in the time series when *Needler* replaced *Cameron*, which suggests that such artifacts did not dominate over other temporal patterns in the data on large-bodied groundfish. (The overlap in areas covered by the Canadian and United States surveys prevents the drawing of parallel conclusions about spatial patterns.) That gives some encouragement to the idea of across-survey comparisons, though it would be rash to assume that a rough equivalency of efficiency among commercial net designs in the capture of resource species indicates a similar equivalency in the capture of the many smaller species (cf. Fowler and Showell 2009) which strongly influence species richness and hence ecological diversity *sensu stricto*.

Thus, in the cases of both of the North Sea and Scotian Shelf trawl survey data sets, we would recommend that mapping and trend monitoring be confined within the spatio-temporal extent across which a single survey protocol has been applied. We further recommend that any differences between the maps or time series arising from different protocols be regarded as consequences of the different filtering and examined in that light, without attempts at standardization or calibration across protocols.

4.1.5 Units of Measurement in Field or Laboratory

Throughout this report, ecological diversity is discussed in terms of relative "abundances". Originally, the required data were literally in units of abundance, meaning numbers of individuals. That makes every sense when the taxocene of interest is confined to adult flying insects, to birds (excluding nestlings) or to adult mammals, since the deterministic growth seen in those taxa leads to a narrow range of adult sizes within any one species. When the idea of ecological diversity came to be applied in forest ecology, however, it was swiftly seen that young oak trees are very much smaller than old ones and hence that "relative abundance" might better be measured in terms of biomass (or, for other taxa, dry weight: Pielou 1975), which more closely represents ecological resource use by the species (Magurran 1988). The same observation does not seem to have arisen as often amongst marine ecologists though indeterminate growth is normal (albeit not universal) in aquatic ecosystems, with most molluscs, crustaceans and fish continuing to grow throughout adult life – though, in contrast to trees, most fish change their trophic relationships as they get bigger: a ton of cod composed of 20 very large individuals is ecologically different from a ton composed of 500 small fish of the same species. Meanwhile, terrestrial botanists have often preferred to use units of percentage cover. They have also faced problems with what have been termed "indiscrete individuals" (cf. Magurran 1988), beginning with those angiosperms that can have very large, continuous "individuals" (e.g. grasses linked by rhizomes) bearing many, individually-small replicate shoots. That too can be a challenge in the sea, particularly with colonial benthic species. The recommended solution is to identify some countable "modular units" and gather data on the relative abundances of those. When measuring the ecological diversity of corals, for example, it might be deemed appropriate to work in terms of polyp abundances (though that would be impractically onerous with most coral species). That approach is limited, however, as there will rarely be any form of module common to more than one major taxon (only the coelenterates having polyps, for example). Other alternatives sometimes recommended (e.g. Magurran 2004) are to use species' range sizes as surrogates for their abundances (though that can only be useful at spatial scales larger than the ranges of the species in question) or else data on "incidence" – the numbers of occurrences of each species in a set of standard sampling units. If values for ecological diversity *sensu stricto* were seen as having ecological meaning, which is difficult to sustain when the relative "abundances" are filtered through some survey protocol, then the ideal might be to work in units of energy use, especially if the taxocene of interest is defined so as to contain only members of one trophic level – though the energy fluxes must usually be inferred rather than measured (McGill 2011). That last alternative might, however, prove difficult to communicate to non-specialist audiences. It is not recommended for use in routine monitoring and reporting.

Any of those measurement units can be used in quantifying species' proportions within a community for the calculation of ecological diversity *sensu stricto*. There is no necessity to represent relative commonness or rarity in number terms (Pielou 1974, 1975, 1977, Magurran 1988; McGill 2011). Whichever alternative is chosen for a particular application, however, it should be used exclusively within any particular diversity estimation. That is: no value of an index of ecological diversity should be built from data on, for example, the numerical abundances of some species and the biomasses of others. To do so would lead to anomalous changes in the index value when biomass-measured species declined and counted species increased, or *vice versa*. Moreover, no value for diversity calculated using data in one measurement unit can ever be validly compared with a value derived from data measured in other units. Not only can numerical abundances not be mixed with biomasses but biomasses measured in grams cannot be mixed with those measured in tons, and so on for the other alternative units.

The parallel question of the use of measures of commonness and rarity in the preparation of Taxon Sampling Curves, for determination of species richness, has rarely been discussed but there does not appear to be any reason why measures other than numerical abundance should not be used, provided that whichever was chosen is used consistently.

4.1.6 Summary

The preceding paragraphs have stressed that, with rare exceptions, ecological diversity can only be determined for one selected taxocene, within the spatio-temporal bounds of a chosen community, and more exactly can only be determined for that taxocene as it is filtered through the protocols of a particular survey. The resulting measurement can then be compared with that of other communities, across either time or space (or perhaps both) but only when the same survey protocols have been used and the same taxocene is considered. That reduces ecological diversity to something very much more constrained than a full quantification of what non-specialists expect of "diversity", even without considering the limitations that ecological diversity has as a facet of biodiversity or environmental status (see Section 2.3).

Where data availability permits, diversity should be determined for more than one taxocene, for more than one season or using more than one survey protocol, since the patterns of variability in time and space that they reveal are likely to be different, while the contrasts may be informative. It should not be forgotten, however, that the ecological meaning of the calculated diversity values will be limited and hence available scientific resources can be better employed in other kinds of analysis, including other analyses of species-abundance data, rather than in multiple determinations of ecological diversity. Expending resources on collecting additional data in order to allow multiple measurements of different ecological diversities would be unwise.

4.2 RICHNESS

Richness (*S*) is the simplest of the three measures of ecological diversity *sensu lato*. Perhaps in consequence, it has been the most studied (while the underlying biology has been explored at length: Rosenzweig 1995), with the additional attention revealing greater levels of complexity than might be anticipated. Indeed, richness turns out to be anything but simple, in either conception or application, and certainly far more complex than merely adding up all species recorded in some area of interest. That is not to suggest that there is anything incorrect in presenting a count of species as an informal descriptor of a fauna or flora, essentially just a summary of a species list, whether that list results from a single sample, the pooled material from a particular survey series or the end result of decades of varied collecting. Such counts can be informative but they should not be confused with values of species richness suited for use in quantitative comparisons among areas or across time. For those purposes, richness must be based on standardized sampling and rigorous estimation methods, lest the comparisons be distorted by artifacts which, particularly in the case of marine ecosystems, could overwhelm any real variations in the numbers of species present.

4.2.1 Curvilinearity in Species Accumulation

One central fact of any field study of richness, and the source of much of the complication, is that the number of species recorded increases with sampling effort but does not do so linearly – a feature of ecological communities that has been recognized since the mid-19th Century (Rosenzweig 1995). Inevitably, the first specimen examined will add the first species to the records. The second and third specimens may well add one or two more species between them but it is very unlikely that the millionth specimen would add a further species to the count, even though several may be added by the second batch of a million specimens. That dependence of recorded *S* on sample size and the curvilinearity in the relationship are so pervasive and so familiar that some of their implications are all too frequently forgotten.

The root of the problem is that, within a chosen taxocene, almost all ecological communities contain a few abundant species, several that are moderately common, plus a large array of rarities. If individuals are sampled at random, extremely large numbers of specimens (mostly of the abundant species) must be examined before each of the rare species is observed even once. More commonly, samples are collected, each containing multiple individuals. It is then extremely improbable that any one sample will include examples of all of the rare species of the taxocene that are present in the community. Hence, it would be necessary to gather a great many samples before being confident of having encountered and recorded every species. Meanwhile, richness weights all species equally, regardless of their abundance, and thus is mostly influenced by the rarities, which outnumber the abundant species in almost all ecosystems (cf. Magurran & Henderson 2011). Hence, large numbers of samples or large sample sizes are essential if richness is to be measured without unacceptable uncertainties.

The ideal solution is to gather sufficient specimens that every species (in the community and the taxocene, as filtered through the selected sampling protocol) has been recorded. To ensure that that has been achieved, an analyst could prepare a "Species Accumulation Curve", in which the total count of species recorded amongst the first 1, 2, 3 ... *N* samples or specimens observed is plotted against the number of samples or specimens. With modern computing capabilities, it is preferable to repeatedly and randomly re-sample the data, thus accumulating the samples or specimens in all possible sequences, not just the order in which they were actually observed. The result is a smooth "Taxon Sampling Curve" (*sensu* Gotelli & Colwell 2001 – termed a "species richness curve" by Hurlbert 1971), accompanied by confidence limits[27]. If such a curve reaches a clear asymptote, then the value of that asymptote (which can be derived statistically, along with its confidence limits) is the species richness of the taxocene of interest in the community in question (Gotelli & Colwell 2001, 2011). For greater clarity, such a result can be distinguished as an "asymptotic *S*" [28].

Values of asymptotic *S* can be validly compared, regardless of the number of samples taken or specimens examined before the asymptote was reached. In essence, once the asymptote has been achieved, the value of *S* would not change no matter how many more samples were collected. Thus, asymptotic *S* can be regarded as the richness that would be observed with infinite survey effort and the values of *S* from different communities compared as though each had received the same, very large amount of effort.

Unfortunately, while it is sometimes straightforward to determine asymptotic *S* (e.g. Rosenzweig *et al.* 2011), as it might be for the extremely depauperate fish

---

[27] It is sometimes claimed that a species accumulation curve carries an implied assumption that the curve is asymptotic (e.g. Gotelli & Colwell 2011), though any such curve must be so if extrapolated far enough, since there is only a finite number of species in each taxocene in the biosphere. It would be more helpful to say that extrapolating a Taxon Sampling Curve to its asymptote carries an implicit assumption that asymptotic *S* is relatively small – dozens or hundreds of species, rather than tens of thousands or more.

[28] Rosenzweig (1995) declared that: "Ordinarily, samples will be complete enough that the simple number of species will be the only diversity index [*sic*: he meant measure of richness] you need or should use." That was simply false: until the asymptote is approached, counts of species are strongly dependent on *N*, as Rosenzweig (1995) himself discussed at length.

faunae of some Canadian lakes, there are many other communities and taxocenes for which that determination is impractical. There are simply too many species of microbiota, plants or invertebrate metazoans, or of tropical vertebrates, for asymptotic S to be determined by any practical amount of sampling (Gotelli & Colwell 2001). The determination is doubtful even for cold-temperate marine fish in some areas. The Scotian Shelf offers a rather extreme example but the teleost ichthyofauna of that area includes literally hundreds of species (Scott & Scott 1988), many of them expatriates carried north from Florida or the Bahamas by the Gulf Stream, while more species continue to be added to the list. Most of the recent additions have probably been deep-living species taken along the continental slope or in shelf-break canyons (e.g. DeVaney *et al.* 2009; Kenchington *et al.* 2009; Halliday *et al.* 2012) but, aside from the deepest of them, those species are not entirely invulnerable to capture by the routine summer groundfish bottom-trawl surveys on the Shelf, which extend onto the upper slope and use nets that fish as they are hauled back through the water column. Should the surveys be continued indefinitely, eventually every fish species that occurs in the area should be taken at least once but, despite hundreds of standard sets per year in a series which has now lasted for upwards of four decades, not all have yet been captured in survey trawls. Shackell and Frank (2003) examined the data from the first three decades of those surveys (in excess of 3,600 standard trawl sets) and found records of only 138 species, plus one genus (*Sebastes*), specimens of which are not routinely identified to their species[29]. The first annual survey, in 1970, took *Sebastes* spp. and 49 other recorded species. Over the next twenty years, further species were added approximately linearly. As of the year 2000, of the 89 species added to the survey records after the first year, 27% had only been taken in one set and 48% in no more than three (Shackell & Frank 2003). Likewise, a survey of fish caught on the intake screens of a nuclear power station in the Bristol Channel (Henderson 2007) found an average of nearly 40 species per year and 65 in ten years but the total number observed continued to increase linearly with additional sampling throughout the duration of the survey, reaching 81 species after 24 years. There was no sign in the data of an approaching asymptote. Of the 81, 14 species were seen only in one year of the series. For each of 10 of those, only a single individual was observed (Henderson 2007). In yet another example, albeit from warmer waters, Chao *et al.* (2009) considered a dataset from trawling in estuarine environments in Georgia (U.S.A.) that had found 70 species amongst nearly 32,000 individuals. Model estimates suggested a total species richness of 90 but also indicated that more than 243,000 specimens would have to be taken before each species was represented in the dataset by at least two specimens (the absence of "singletons" being taken as evidence that no other species were present: Chao *et al.* 2009).

Even when some set of species-abundance data is sufficient to generate two comparable values of asymptotic S (corresponding to two areas at the same point in time or else to the same area over two time periods), there can be strong incentives to subdivide the data more finely. That is especially true when mapping spatial variations in richness, since a map with a spatial resolution coarser than that of the management questions to be addressed would be next to useless. Yet only rarely will there be sufficient data to determine asymptotic S for every square of a fine map grid or even annual asymptotic S for a survey series.

---

[29] Shackell and Frank (2003) claimed 140 "species" but counted "*Notolepis rissoi*" and "*Notolepis rissoi kroyeri*" (both synonyms of *Arctozenus risso*) as two.

Hence, whether because of a very large number of rare species or because of a desire to more-finely divide the available data, it is commonly necessary to work with counts of recorded species within the chosen community or communities, rather than with asymptotic richnesses. Yet, those recorded counts will vary not only with the community richnesses themselves but also with the amount of sampling effort devoted to each time period or community. If additional sampling were directed to an area of conservation interest, its recorded counts of species would rise, while reduced budgets for long-term monitoring would lead to temporal declines in recorded "richness". Clearly, neither outcome would be satisfactory.

Ecologists have long represented these relationships in the form of "species-area curves", in which the number of species is plotted against the area surveyed, revealing a ubiquitous curvilinear relationship, heading towards an asymptote at the number of species in the community, meaning asymptotic S[30]. In recent years, it has become clear that there are also similarly-shaped species-time curves and there have been calls for development of species-time-area relationships (Magurran 2011). Two, or by some counts four (e.g. Rosenzweig 1995), different processes contribute to those curves, often with more than one acting within the same dataset. One is a refined version of the sampling issue that has already been noted: drawing larger samples from within some homogeneous habitat patch involves extracting specimens either from a larger area (e.g. sweeping more seabed by deploying more standard trawl sets), from the same area over a longer a period of time (e.g. maintaining deployed gillnets at the same fixed stations for longer periods) or a combination of the two. The increased number of specimens leads to more rarities being recorded but the increase in species count will not be linear with survey effort – not, at least, after some initial phase that suffices to record all of the reasonably abundant species. Secondly, enlarging the defined spatial extent of a community will inevitably broaden the variety of habitats within its boundaries and will thus tend to increase the total community species richness, which almost always means higher recorded species counts from the same survey effort (Rosenzweig 1995; Rosenzweig *et al.* 2011)[31]. The mechanisms driving that broadening among areas of varied extent arbitrarily delimited within some larger region differ from those which cause the similar phenomenon among ecological "islands" (which are not necessarily patches of land surrounded by water) of varied sizes (Rosenzweig 1995). Likewise, aside from "closed" communities that cannot receive invasives from outside (e.g. some glacial lakes, at least for some taxocenes), extending the duration of a community's temporal bounds will allow time for some species turnover, tending to increase the total number that occur within those bounds. For richnesses that respond to diel or seasonal cycles, even extending a survey from daylight-only sampling to a full 24 hours or from summer-only recording to a year-round operation may massively increase the species count within the overall temporal bounds of the community. The two processes, increased sampling within fixed spatio-temporal bounds or broadened bounds, should be clearly distinguished (Magurran 2004) but are not always.

---

[30] A species-area curve is similar to a species-accumulation curve but they are only identical if the independent variable of the accumulation curve is plotted as equal-area units of survey effort.

[31] As the extent of the community under study is increased, it could encompass patches containing quite different species, such as when a neritic study expands to include portions of the continental slope. (Something similar can occur in very large scale studies if the increase in extent leads to a second biogeographic province being included in the data collection: Rosenzweig 1995.) In such cases, a species-area curve can reverse, steepening with additional sampling

It follows from this curvilinearity that, if small spatial variations in richness are to be mapped or slow temporal change in richness detected in situations where asymptotic *S* is unavailable and if the variations or the change are not to be overwhelmed by artifacts, it is essential that comparisons be made among communities with comparable spatial and temporal bounds. There may be questions in conservation planning that can only be answered if the richness of some large, offshore bank is compared with that of a much smaller, inshore feature, the bathymetric reality of each feature giving validity to their comparison. More often, mapping needs richness determinations for equal-sized areas, while temporal trends must be examined using data binned across equal time periods. In particular, it would usually be a mistake to map richnesses following the unequal-area strata that are typical of resource surveys. Moreover, should an attempt be made to use finer spatial grids when mapping areas of greater conservation interest, and coarser ones in surrounding areas, there is likely to be a perverse reduction in apparent richness wherever cell size is smaller.

In addition to equal-area and equal-duration communities, if values of asymptotic *S* cannot be determined, valid comparisons among richnesses demand that the same sample size (meaning, in most applications, the same number of individuals, summed across species) be taken from each map area or each time period, lest recorded species counts be driven by the artifacts of survey planning rather than by biology. Since real sample sizes will almost always differ from community to community, it is usually necessary to standardize the data to some chosen size. The appropriate technique is rarefaction (see Section 4.2.4 below) but too many ecologists have sought simpler standardization using ratios. Unfortunately, the curvilinearity in species-time-area relationships precludes that approach. While an area of 1,000 km^2 will almost always contain more species of the taxocene of interest than a nearby area of 10 km^2, it will very rarely contain 100 times as many. Just as the raw species count in the larger area would be biased upwards, so calculated values expressed as species per 1 km^2 would be

---

before heading for a new and higher asymptote. That reversal would be obscured if the curve is drawn using repeated random re-sampling of the data.

biased in the opposite direction. The same could be said of extending the duration of a survey or of increased survey effort. Quiet simply: Ratio measures fail when applied to non-linear relationships (Rosenzweig 1995; Gotelli & Colwell 2001). Early quantitative ecologists sought to avoid that problem by using ratios of species count to some non-linear function of sample size. Two examples of that approach, Margalef's and Menhinick's indices, continue to appear in textbooks (e.g. Magurran 2004) and sometimes in research studies (e.g. Greenstreet & Rogers 2006). Those indices, however, assume that the number of observed species increases with, respectively, the logarithm or the root of sample size, when the actual relationship is strongly dependent on evenness. Hence, while either may work well enough with a few datasets, they risk returning grossly misleading values if applied to other data.

An alternative attempt at simplification seeks to apply equal survey effort to each community or at least to standardize by survey effort rather than sample size. That leads to a subtle but important dichotomy. In the common parlance of ecologists and conservation biologists, the number of different species in any unit is said to be its "richness", whether that unit is a single sample, a particular habitat patch, a single cell on a map grid or something else. More precise definitions restrict the term "species richness" (alternatively: "numerical species richness": Hurlbert 1971) to the number of species recorded in a sample containing a fixed number of individuals (or a fixed biomass), while the count of species in a certain area or other unit, such as unit sampling effort, is considered to be its "species density" (Magurran 1988, 2004; Gotelli & Colwell 2001, 2011) – an unfortunate choice of term as the curvilinearity in species / area relationships precludes expressing species density in units of "species per unit area" and it is always necessary to state the area across which a species density has been determined. The dichotomy between species richness and species density arises because numbers of individuals (summed across species within a taxocene) vary among areas and, because of variations in density, they do so even if the extent of each area is fixed. Abundances likewise vary over time. Since observed species richness varies with the number of individuals examined, species density is a function of both abundance and richness – or it is so unless sufficient samples have been collected for asymptotic *S* to be determined[32]. Short of that achievement, species density is a poor measure of ecological diversity (Gotelli & Colwell 2001), though many published reports of "species richnesses" actually present species densities.

This difference between species richness and species density may appear too subtle to merit much attention but it is commonly seen that they yield reversed relative rankings of communities. In one experiment on the effects of disturbance on freshwater macrobenthos, a decline in "richness", which was both expected and observed, proved to actually be a decline in abundance and hence in species density, while species richness *sensu stricto* remained unchanged (Gotelli & Colwell 2001). When tracking temporal change (especially in the open sea, where extirpations are rare but major changes in abundances are common) and where insufficient sampling has been completed for asymptotic *S* to be determined, numbers of species should always be expressed in terms of "numerical" species richness. In contrast, when mapping diversity for conservation planning purposes, species densities are often the relevant measure: If the aim is to protect the most species within the smallest closure, then it is the count of species per unit area which matters, without regard to whether or not that is driven by higher abundance (Gotelli & Colwell 2001).

---

[32] Counts of observed species may ascend towards the asymptote at different rates, depending on whether they are accumulated by units of survey effort or by the number of individuals examined but either must yield the same asymptotic *S* if sampling effort and sample size were high enough.

A further complication arises that Gotelli and Colwell (2001) did not address and which should be outlined here: In temporal trend monitoring, if the standardization to a fixed sample abundance is correctly computed, the standardized count of species will be independent of the inter-annual variations in the number of individuals taken in a survey and will therefore be a "numerical species richness" (though not the community's asymptotic *S*). It will be independent of variations in survey effort and also of those in survey catch rate, the latter presumably approximately proportional to the abundance of individuals in the taxocene, within the community. However, the standardized richness will not be independent of changes in the evenness of the community. Should the overall abundance of the taxocene fall because the abundant species have been depleted, while the numbers of individuals of the rare species remain steady, as would be the expected effect of fishery depletion, then fewer of the individuals in the standard-sized sample will come from abundant species, leaving more "slots" to be filled by representatives of rare species and hence a higher count of species. Thus, unless asymptotic *S* can be determined, measured species richness *sensu stricto* is related to evenness and is capable of displaying temporal trends that are artifacts of its calculation, just as species density is.

Declining abundances of rare species can also create artifacts, if the numbers of the abundant species are maintained. In time, each declining rarity would eventually reach an abundance at which it is only infrequently taken by a particular survey and would usually do so long before it being extirpated from the monitored community (because of low, but unknown and variable selectivities). Counting the species in a fixed sample size, rather than determining asymptotic *S*, will mean that some rarities are included in the count in each time period (e.g. each annual survey) while most are not, with the number that go unrecorded increasing as abundances decline. Thus, the trend in observed richness will reflect not just *S* itself but also the (absolute) abundances of the rare species. In the open sea, where extirpations are rare and global extinctions of species essentially unknown, the trend will be almost exclusively one in abundances, yet it will too readily be labelled as a change in richness. Only when all species increase or decline in abundance while maintaining their same relative abundances will a species richness, properly determined from a fixed number of specimens, track the number of species in the community and taxocene (as the latter is filtered through the survey protocol). Thus, where it is impossible to determine asymptotic *S*, time series of richness should never be considered in isolation from evenness. Determining evenness without first obtaining a value for asymptotic *S* is, however, problematic – as explained below (see Section 4.5).

4.2.2 Determining Sample Species Richness and Species Density

Before community richness can be determined, it might be desirable to determine sample species richness or (more often) species density for each of a large number of samples. The richness or density is simply the number of discrete species (in the taxocene of interest) in the sample – the result being a species density if the sample resulted from unit sampling effort or the census of a unit area but a species richness if the sample contained a standard number of individuals.

If every individual has been correctly and fully identified, determining the number of species is trivially easy. Some care is needed with many datasets, however, since it is common to find that a proportion of individuals have only been identified to genus (or even some higher taxon). Those may be incompletely identified (perhaps damaged) members of a species already recorded as present or their identification to genus may be an indication that they are members of unfamiliar species, not otherwise recorded in the dataset. When working with poorly-studied communities, such as the benthos of oceanic depths, multiple species new to science may be encountered and field survey teams can differ in how those are designated. Moreover, taxonomic revisions can lead to nomenclatural changes during long survey series, with no certainty that the older data have been corrected to match modern names. Indeed, it may not be possible to revise the early data if the revisions involve splitting of species, long after specimens have been discarded at sea. All such complications are usually easily resolved when the field data were gathered under the diversity-analyst's control but use of archived data may necessitate careful attention to the protocols used for the naming of specimens during former survey programs.

If those challenges can be met, leading to an unambiguous count of species taken by unit sampling effort, then determining sample species density is straightforward. Being derived from a complete census of the sample, it is known exactly and has no variance.

Some survey protocols select individuals at random and cease when a sample of a certain sample size has been achieved, allowing a species richness to be determined equally easily. Given more typical survey methods, such as trawling or the deployment of a benthic grab, a fixed number of specimens could be randomly selected from each catch. That randomization is, however, technically demanding, while there is usually a desire to gather data on the entire catch. Thus, available survey data do not normally lend themselves to determination of sample richnesses by straightforward counting of species. Should sample species richnesses be needed when sample sizes are unequal, it would be necessary to resort to the rarefaction techniques discussed below (see Section 4.2.4 below) and the richness then determined for some standard number of individuals. Whether that effort would ever be justified must be doubtful since all that would be determined would be sample richness and the analytical effort might be better applied to estimating the species richness of the community directly.

In that context, it perhaps bears emphasizing that either the richness or the species density of a sample is of little ecological or management relevance, since either is but a point value which almost always grossly underestimates both the richness and the species density of the community from which the sample was drawn, while being subject to very considerable sampling error. Except in those rare cases where it is possible to arrange the spatio-temporal scale of each sample to match the scale that is relevant to the management (or research) issue at hand, sample richness should be seen as an α diversity, whereas the measurement required is a community- or γ diversity across the relevant scale. Since β diversities cannot be estimated from α diversities alone, the latter are rarely useful other than as a step in determining the community γ diversity.

4.2.3 Community Species Richness and Species Density

While the determination of sample richness and species density is not without its challenges, much larger complications arise when samples are combined in the determination of the richness or species density of a larger community. It should not need to be said that the sample richnesses cannot be summed (since many of the species will be common to multiple samples), nor averaged (since the richness of the community must at least equal, and will almost certainly exceed, that of the richest sample). Pooling the data from all samples drawn from the community and then determining S as the number of discrete species would come closer to the correct result but is still not sufficient[33].

To properly determine either richness or species density, a Taxon Sampling Curve must be prepared (see Section 4.2.1). If that curve reaches its asymptote, allowing a firm determination of asymptotic S, then the independent variable may be either of the number of accumulated specimens or units of survey effort, whichever is more convenient[34]. If the asymptote cannot be achieved with the available data, the number of accumulated standard units of sampling effort is the appropriate independent variable if (but only if) the relevant measure, for the question at hand, is species density. Otherwise, meaning whenever the required measure is species richness and asymptotic *S* cannot be determined, the number of observed species must be plotted against the accumulated number of individual specimens. In those surveys that allow random selection of individual organisms (as is sometimes possible in forest ecology), the individuals should be accumulated one by one. With the more typical form of survey, in which collections of individuals are taken together (each collection by one unit of survey effort), Taxon Sampling Curves should still be drawn with the accumulated number of individual specimens as the independent variable but they should be accumulated collection by collection, rather than individual by individual. Accumulating the data in that way incorporates the among-effort-unit patchiness in the communities into their Taxon Sampling Curves (Gotelli & Colwell 2001).

---

[33] A particular caution is needed here with respect to geostatistical approaches to data analysis. Those approaches include powerful tools for drawing two-dimensional representations of sets of geo-referenced point data, with immense value in ecology. Some of those tools have been used in mapping species richnesses (e.g. Horsman *et al.* 2011; McGill 2011). Great care is needed in doing so, however, since few other data types involve the sorts of curvilinearity inherent to counts of species and it cannot merely be assumed that geostatisticians considered the suitability of their techniques for such data. Diversity analysts should examine the mathematical structures of any geostatistical tools that they consider using to ensure that those are compatible with the particular applications intended. Given the frequency of errors in the literature, the fact that some analytical tool has been commonly applied to diversity data in the past cannot be sufficient evidence that it is appropriate for that purpose.

It should be noted that this requirement to draw Taxon Sampling Curves against an independent variable defined in terms of numbers of specimens disproves the oft-repeated claim that species richness can be determined from presence / absence data alone, making it simpler to apply than is ecological diversity *sensu stricto*. A count of all species (in the chosen taxocene, as filtered through the selected survey protocol) present within the defined bounds of some community cannot be accepted as species richness because it cannot be known whether that value is asymptotic *S*. Lists of species in each of many sample collections are sufficient to prepare a Taxon Sampling Curve (provided that each collection was taken using the same survey protocol) and should that reach an asymptote, or should the interest be in species density, no more complex data are needed. However, in the common situation that species richness must be determined from a Curve that does not reach an asymptote, either the number of individuals examined must be counted or else the number in each sample must be known, so that they can be summed into a number of accumulated specimens.

Strictly speaking, preparation of a Taxon Sampling Curve assumes that the data were randomly collected from the community under study, sometimes with each specimen being selected at random but, more often, with randomly collected samples and a complete census of all individuals (of the taxocene of interest) in each sample (Gotelli & Colwell 2011). Stratified-random or uniform sampling appear to be adequate alternatives in most cases. Opportunistic sampling or sampling focused on some other topic of research interest (and merely utilized for estimating richness) will typically lead to underestimates of the community's asymptotic *S*, since either will tend towards listing all species in some subset of the community, while under-representing those that live elsewhere. Conversely, survey effort directed towards capturing rarities would tend to alter the curvature of the Taxon Sampling Curve, leading to mis-estimation when it is necessary to extrapolate to an asymptote and uncertain behaviour when the analyst resorts to rarefaction. Non-randomness can also violate the assumptions of some statistical techniques designed for use with species-abundance data (Gotelli & Colwell 2011). Thus, in some applications, it may be necessary to narrow the dataset, the defined bounds for the community of interest and / or the taxocene considered in order to achieve near-randomness. If that is done, care may be needed to avoid estimating an *S* that is irrelevant to the management questions handoff immediate concern. Most certainly, the data must be gathered through some systematic survey, rather than haphazard accumulation of records (Magurran 2004). The latter would suffice if merely counting numbers of species known to be in a community but determining richness in terms that allow quantitative comparisons across time and space is more demanding. Systematic data collection is especially important when extrapolating or rarefying Taxon Sampling Curves as either process assumes some consistency in the behaviour of the data, which requires that the data collection was itself consistent.

---

[34] In most marine applications, the data are well grouped by units of survey effort (e.g. standard survey trawl sets, benthic grab drops) and a sample-based curve is convenient. However, some datasets do not lend themselves to analysis in terms of such units. For example, successful hand-lining for fish depends strongly on whether the fish are biting at the time and location of the sampling, such that "hand-line hours" may not be a useful measure of survey effort. Much the same could be said of electrofishing in freshwater streams. For those, Taxon Sampling Curves may more easily be drawn with the accumulated number of specimens as the independent variable.

When a Taxon Sampling Curve fails to reach its asymptote, it is sometimes possible to extrapolate to an estimate of asymptotic *S*, along with confidence limits around the estimate. However, the process of extrapolation requires an assumption of some functional form (e.g. the negative exponential or the familiar equation of Michaelis-Menton kinetics) that can be fitted to the data. Unfortunately, it is commonly seen that all available functions fit equally badly and yet predict very different asymptotic *S*, resulting in excessively high uncertainties in the extrapolations (Gotelli & Colwell 2011). Given data on the relative abundances of the species, and if a functional form can be assumed for the Species Abundance Distribution, an alternative way to estimate asymptotic *S* is to fit a curve to the SAD and read off the number of species with zero abundance. Adding that number to the count of observed species should provide an estimate of richness. In practice, however, that approach works no better than extrapolating the species accumulation curve (Gotelli & Colwell 2011). Graphical approaches have been suggested as another alternative but they provide no advantages in an era of abundant computing power and available packaged software[35].

---

[35] Karakassis (1995), for example, suggested plotting the number of species observed in $v+1$ standard samples against the number observed in v, with those numbers drawn from averages of multiple permutations of the sample sequence. Asymptotic species density, equal to *S*, can be read where the fitted line cuts a 45° line on the graph. Karakassis (1995) found that that approach worked well with two datasets on marine benthos, though he did not offer confidence intervals around estimated *S*. The shallow angle of cut between the two lines drawn for each example suggests that those would be wide. Moreover, and Karakassis' (1995) claim to the contrary

It is possible to estimate a lower bound on S by adding to the count of recorded species an estimate of the number of as-yet undetected species (in the community and the taxocene as filtered through the survey protocol) that is derived from the ratio of the number of "singletons" to the number of "doubletons" – respectively the species represented in the dataset by one or two individuals each. Of the available estimators, the two most promising (Gotelli & Colwell 2011) are "Chao1", applicable to species abundance data on a random collection of individuals, and "Chao2", for presence / absence data based on randomly distributed collections (e.g. quadrats, trawl sets or grab drops). The estimators are:

$$Chao1 \quad \cdots \quad S_{est} = S_{obs} + {f_1^2}\Big/{2f_2}$$

$$Chao2 \quad \cdots \quad S_{est} = S_{obs} + \left(1 - \frac{1}{t}\right) {Q_1^2}\Big/{2Q_2}$$

where S_{est} is the estimated lower bound on asymptotic S, S_{obs} is the number of species observed in the dataset, f_r is the number of species represented in the dataset by r specimens, and Q_r is the number of species recorded as present in r samples of a total of t samples. Corresponding variance estimators are available, as are estimators for cases where f_2 or Q_2 is zero (Gotelli & Colwell 2011). The approach is only fully applicable to homogeneous data and is apt to yield errors if the sampling was spread across an ecological gradient, since that would distort the relationship between doubletons, singletons and undetected species (Magurran 2004). The "Chao1" estimator is an approximation only applicable to datasets containing large numbers of individuals but that is normal for any set of species-abundance data used in estimating ecological diversity. More seriously, the derivations of both estimators ignore the rather complex statistics of rare events, such as the recording of a very rare species. Nevertheless, they seem to perform well when applied to empirical data, "Chao1" even performing well with data gathered from random survey units instead of the random selection of individuals assumed in its derivation (Chao *et al.* 2009).

The principal weakness of the Chao estimators is that they only provide estimates of a lower bound on S. Trials show that, as expected, the bound increases towards an asymptote equal to asymptotic S as sample sizes increase but also that it approaches that value with much less survey effort than observed S does (Chao *et al.* 2009; Gotelli & Colwell 2011) – except that, when sample sizes are very small, Chao estimates are highly imprecise and can exceed the asymptote seen with larger sample sizes (Magurran 2004). Whether the S_{est} calculated from a particular dataset is an adequately close estimate of asymptotic

---

notwithstanding, any such graphical approach involves an assumption of the functional form of the extrapolation to the asymptote, that assumption being embedded in the mathematical transformation that linearizes the Taxon Sampling Curve. In Karakassis (1995) particular method, the negative exponential is assumed. That evidently fitted his trial datasets well but will not be applicable to all others.

S is, however, something that must be confirmed on a case-by-case basis. Either "Chao1", if abundance data are available, or "Chao2", if the data are only presence / absence, should be applied to variously-sized, randomly-chosen subsets of the data and the resulting estimates plotted against subset size, creating a graph analogous to a Taxon Sampling Curve. Only if that forms an asymptote, or can be extrapolated to one without unacceptable uncertainty, can the value of S_{est} be taken as an estimate of asymptotic *S* (cf. Gotelli & Colwell 2001)[36].

4.2.4 Rarefaction

When it proves impossible to develop an adequately precise value for asymptotic richness by any of the above approaches, it is necessary to resort to rarefaction. That process uses the same Taxon Sampling Curve but instead of extrapolating towards the asymptote, rarefaction interpolates to some chosen, and smaller, sample size. By rarefying the data from multiple communities to the same sample size, comparisons amongst them are freed of the effects of variations in survey effort and abundance-density, without the use of invalid ratio approaches (Magurran 1988; Gotelli & Colwell 2001, 2011). In effect, rarefaction uses the data actually collected to determine what the observed species count would have been if data collection had been cut off at some smaller, but standard, sample size. The subsequent comparison of rarefied estimates of *S* is valid in the sense of comparing like with like (counts of species derived from the same sample size). Whether it bears any relationship to a comparison of the communities under study is quite another matter.

When the aim is to monitor trends in species richness over time, the data from each temporal community should be used to develop a Taxon Sampling Curve (one per community) with the accumulated number of individuals as the independent variable. If the data came from a survey that collected random individual specimens, the latter should be selected at random when constructing the curve. Data from complete censuses of the biota taken by randomly-distributed units of sampling effort (e.g. from the catches taken by trawl sets) should be accumulated by units of sampling effort (e.g. set-by-set) in Curve construction, though the number of observed species should still be plotted against the number of specimens, not that of units of effort, as explained above. A standard sample size (in numbers of individuals) should then be chosen, typically as the smallest number recorded in any one temporal community or else as a convenient round number somewhat smaller still. For each time period, the rarefied *S* corresponding to the chosen sample size, along with confidence intervals around it, can then be read off the graph for that time period or, in practice, can be calculated by appropriate software.

---

[36] Asymptotic values of the Chao estimators may not always be such reliable indications of asymptotic *S* as Gotelli and Colwell (2001) suggested. Shackell and Frank (2003), for example, applied both Chao1 and Chao2 to data from the Scotian Shelf groundfish trawl surveys conducted from 1970 to 2000. Both reached asymptotic values with less than 30 years' data. The resulting estimates of richness were however 172±18 (Chao1) and 164±12 (Chao2), when the known ichthyofauna of the region contains at least twice as many species. The causes of that underestimation have not been investigated but might involve heterogeneities within the surveyed area – many of the species present being vulnerable to only a few of the sets made each year, particularly those few in deep water on the upper continental slope.

When, in contrast, the aim is to map variations in species density, the process is almost the same but uses Taxon Sampling Curves (one per spatial unit in the map) with standard units of survey effort (e.g. trawl sets) as the independent variable, while the observed species densities will usually be rarefied to the least number of units of survey effort applied in any one spatial unit[37].

Despite its widespread use, rarefaction has been roundly criticised, particularly for its wasteful discarding of data from times or areas that had seen plentiful sampling (e.g. Magurran 1988). It also moves the basis of comparisons away from the portion of a Taxon Sampling Curve that contains most information on rarities, meaning the vicinity of the asymptote, and towards the data-poor region, where underestimation of the true community *S* is greatest and where all such Curves converge onto the point at which a sample of one individual necessarily contains one species. In the process, the statistical power to discriminate amongst communities is discarded (Gotelli & Colwell 2001, 2011; Rosenzweig *et al.* 2011). Even when rarefaction does not proceed quite so far, it can be badly distorted by the clumping of individuals which is a common property of many species but can be especially marked in some kinds of benthos: any one grab sample might contain dozens of individuals but most of those might be members of one species per sample, depressing the Taxon Sampling Curve (Gray 2000). A yet more serious deficiency is that different communities show different trends in recorded species count with increasing sampling. The shape of an individual-based Taxon Sampling Curve depends on the evenness of the community: higher evenness giving a steeper curve as the first several specimens are likely to be members of different, equally abundant species. When evenness is low, only the few abundant species are likely to be recorded amongst the first specimens, while the many rarities are added slowly as much larger samples are accumulated (Magurran 2004). In the case of the sample-based curves used in rarefaction of species densities, it is patchiness that determines the shape differences (Gotelli & Colwell 2001). If, hypothetically, marine benthos was sampled by grab, with one study area being on the top of a broad, sandy offshore bank while the other was an inshore patchy area, with a mix of seabed types extending from cobbles to fine mud, it would not require much offshore sampling to list all of the common species of the area and a share of the less-common ones. In the inshore area, however, achieving the same quality of information would need as much sampling effort put into each of the types of patch as the total needed on the bank. Since communities differ in both their asymptotic richnesses (or species densities) and the rate at which those are approached by their Taxon Sampling Curves, it follows that two Curves may cross and indeed they sometimes cross twice. Hence, if the species richnesses of one community at various times are compared (as in trend monitoring) or if the species densities in the cells of a map grid are compared (as in spatial planning for conservation purposes), the relative rankings of the time periods (meaning the temporal trend) or those of the mapped areas can depend on what sample size is chosen as the standard to which all others are rarefied (Hurlbert 1971; Peet 1974; Gotelli & Colwell 2001). In short, conclusions about whether richness is increasing or decreasing, as well as those about which areas have the higher species density, can be nothing more than artifacts driven by the intersection of contrasting curvatures in Taxon Sampling Curves and the amount of survey effort devoted to the most data-poor temporal or spatial community in the study. Rarefied species richnesses and species densities need bear no relationship to the asymptotic values for the same communities, not even necessarily maintaining the same rank order, and thus may have no ecological significance whatsoever.

---

[37] Some authors confine the term "rarefaction" to the use of curves drawn against an independent variable scaled as individual specimens and hence yielding species richnesses. Curves drawn against survey effort and used to derive species densities are then called "randomized" or "smoothed" accumulation curves (e.g. Gotelli & Colwell 2001).

With such grave deficiencies, it must be questionable whether rarefaction merits the effort required for its calculation. Rosenzweig *et al.* (2011) were particularly condemnatory, declaring that "rarefaction does not so much reduce sample-size bias as surrender to it" and that the comparisons made among rarefied samples "have little to do with the diversity signal". Nevertheless, there is no better alternative when the data are insufficient for estimation of the asymptote of a Taxon Sampling Curve. Perhaps the best advice would be to use asymptotic *S* where that can be estimated with adequate precision, to resort to rarefaction only when the smallest sample size is large, though not large enough to allow estimation of asymptotic *S*, and to refuse to report species richness or species density at all when the sample size to which the data would have to be rarefied is so small that differences among communities in rarefied *S* cannot reasonably be taken as indicative of their differences in asymptotic *S*. The proper response to such deficient sample sizes is not a comparison among times or places that, while numerically valid, would be divorced from ecological meaning. Rather, what is needed is either pooling the available data into larger units (e.g. decadal, rather than annual, for temporal comparisons; 1000 km^2 cells, rather than 100 km^2, for mapping) or else undertaking additional fieldwork to collect the data required for meaningful comparisons among the units of management interest. The adequacy of rarefied *S* as an indicator of asymptotic *S* might be judged by examination of the Taxon Sampling Curves: if their curvature indicates that larger samples would not have led any of them to cross, then at least the rarefied values preserve the rank order of the asymptotic ones. Conversely, if rarefaction cuts off two or more sample-rich curves below a point at which they cross, then the rarefied *S* are essentially meaningless.

One consequence of that advice is that spatial variations in species density cannot meaningfully be mapped using such a fine grid that only a few samples have been taken from each cell, unless the selected taxocene is so depauperate within that map unit that even the few samples suffice to provide a value for asymptotic *S*. Otherwise, the grid must be coarse enough that even the most data-poor cell still has enough data to support useful estimation of local rarefied *S* (strictly: species density). Whether any dataset held by the Department is extensive enough to allow that estimation for the cells of a grid that would be fine enough to be useful in spatial planning is a question that can only be answered during attempts to produce the maps. If existing datasets are not adequate, the only acceptable response is to refrain from mapping until more extensive data become available. Generating maps of rarefied *S* that have a different rank order to the asymptotic *S* for the same grid cells would only serve to direct conservation efforts to the wrong places. Depending on how the data are mapped, even the correct rank order may not suffice.

Hurlbert (1971) took a different approach, recommending that comparisons of richnesses always be made by examination of the Taxon Sampling Curves themselves – a position echoed by Gotelli and Colwell (2001). The latter described attempts at comparisons without considering the Curves to be "problematic at best". That was no doubt good advice for application in ecological research, in which the shape of the curves can convey information on patchiness as well as indications of richness. However, to turn from species richness to a Taxon Sampling Curve is to accept that a scalar index value is insufficient to represent the ecology of the communities of interest and hence to reject the usefulness of measures of ecological diversity. Certainly, routine monitoring of temporal trends cannot usefully be reported as a set of overlain curves, with the non-specialist user invited to judge whether richness is increasing or decreasing from their examination. Nor can a map be drawn with a curve for each grid cell.

4.2.5 Some Common Errors

As the preceding paragraphs make clear, for most communities and taxocenes, properly determining a value for species richness or species density is not the simple task that it is often represented to be. It is onerous and commonly very demanding of expensive field data, those demands generally preventing fine-scale mapping or the tracking of temporal change without binning data into decadal, or longer, pools. Many authors have claimed to evaluate richness without confronting those challenges but too often they have substituted sample richness for community richness, or species density for species richness.

The former error has been seen in attempts to map spatial patterns in species densities (often mistaken for richnesses) derived from trawl-survey data and to map them at scales that permit only one, or at most a few, survey sets per cell in the map grid – sometimes using mean sample densities instead of resorting to rarefaction (e.g. Shackell & Frank 2003; Horsman *et al.* 2011; ICES 2011b). Those attempts have been prevalent enough in marine applications to merit consideration of what the outputs mean.

Typical trawl surveys were designed to provide estimates of trawlable biomasses, meaning (for each of a number of species) the product of biomass, within some survey frame, and the selectivity of that species under the survey protocol. If the sampling effort could be increased until every square metre of seabed within the survey frame was swept once and only once by the survey gear (and, strictly speaking, if all of those sets could be made simultaneously), then the sum of the catches in the sets would be the trawlable biomass and would equal the mean of those catches multiplied by the number of standard sets needed to fully sample the survey frame. It follows that the mean of the catches taken in the sets of a real survey, multiplied by the same areal-expansion ratio, provides an estimate of the trawlable biomass of the species in question. In contrast, when working with species richness or species diversity (or with ecological diversity *sensu stricto*), the curvilinearity in species-area relationships means that the sum of the richnesses in enough sets to cover the entire survey frame would vastly exceed the richness of the area: each ubiquitous species would be counted thousands of times, while even species so rare as only to be recorded twice would be double counted (unless both specimens were taken in the same set). Since summation of richnesses or species densities leads to such absurd results, means are not useful guides either.

What a sample species density may represent is the α diversity within the area swept by a single trawl set –about 40,000 m^2 or 0.04 km^2 in the case of the Scotian Shelf surveys (cf. Shackell & Frank 2003)– and the temporal duration of that set. (Many rarely-caught species, whether actually rare or having low selectivities to the survey gear, will pass through the area swept in the course of a decade but not be taken by a specific set.) The mean of per-set densities is thus the mean of such α diversities. Management concerns in the offshore areas where trawl surveys are conducted are, however, likely to focus on scales of 1,000 to 10,000 km^2 –approximately the typical size of offshore MPAs or other management zones– and durations of at least a year if not decades. What is required of an estimate of ecological diversity for most applied purposes is the γ diversity of a community defined at some such larger spatial and temporal scale – and that cannot be found from the α diversities of the samples by summation, averaging or even by pooling but needs the more complex approaches outlined above.

These are not minor concerns, since β diversity is so variable that γ cannot be predicted from average α diversity. Consider a hypothetical case in which the eastern half of some survey area has a richness of 50 fish species, ten of them commonly encountered and the others rare enough that an average set takes one in eight of them. Suppose that that portion of the survey area is spatially homogenous on large scales, with its patchiness largely expressed at scales of 10,000 m^2 or less (i.e. smaller than the area swept by a single trawl set). Typical survey sets would take the ten common species and five of the rarities, for an average species density per set of fifteen. Suppose that, in contrast, the western half of the survey area contains three quite different habitat types, intermingled on scales of 10 km^2 (i.e. large enough that most trawl sets fall within one patch but much smaller than one half the entire survey area). Because each patch is relatively small, it would be expected to have lower richness than is seen in the eastern portion of the survey area but because the western half is more variable, it would be expected to have higher richness overall. If each habitat type held six common species and 24 rarities, one in eight being taken by a typical survey set, then sample species densities would average nine, yet the entire richness of the western portion of the survey area would total 90 species. Mapping mean sample species densities would then suggest that "richness" is nearly twice as high in the east (fifteen *versus* nine) when richness on the scale of MPA designs would really be nearly twice as high in the west (90 *versus* 50). In that hypothetical sea area, an MPA designed to protect as many species as possible within a given closure extent should be placed in the west but a map of average sample species densities would indicate that it should be placed in the east – exactly the wrong outcome.

Yet that erroneous use of mean sample diversities is commonly encountered in the literature and not only when trawl-survey data is being mapped. Callaway *et al.* (2002), for example, claimed higher epibenthic species richness in the northern North Sea than in its Southern Bight based on averages of sample species densities. Schratzberger *et al.* (2006) sought to understand patterns of nematode diversity in the North Sea using α diversities on the scale of 35 cm^3 (3.5×10^{-5} m^3) sediment samples and averages of trios of those. Other authors have simply used single sample diversities (e.g. Jennings *et al.* 1999; Rachor *et al.* 2007), which at least have the advantage of being measured over a known sampling scale, rather than averaging diversity among samples.

In a different variant, Greenstreet and Rogers (2006) concluded that S declined through the 20th Century in heavily-fished portions of the northwestern North Sea, carrying ^1D and ^2D down with it. Unfortunately, their measure of S was not the asymptotic richness in some large and long-term community (i.e. a γ diversity on scales of management relevance) but rather the species density in a pool of ten standard survey sets (an α diversity). In addition to the problems of spatio-temporal scale encountered in the mapping examples, the use of species density in place of richness when examining temporal change made Greenstreet and Rogers' (2006) "S" sensitive to abundance. As fish populations declined, under fishing pressure, species density would have declined even without any change in richness[38]. Greenstreet and Piet (2008) subsequently estimated that a minimum of about 20 sets of the North Sea International Bottom Trawl Survey must be aggregated before calculating the ^0D, ^1D or ^2D metrics. That number was not, however, suggested as being sufficient to determine asymptotic values for any if the three indices, and clearly was not even close to being adequate for finding asymptotic *S*. Greenstreet and Piet (2008) did suggest that relative rankings of the asymptotic values could be adequately determined with just 20 sets if the dataset were reduced to "members of the 'demersal fish community'", a step which likely eliminated those species that are rare in the full dataset. That was an effective way to reduce data requirements and thus allow mapping but it carries the twin risks of generating artifacts, through arbitrary allocation of species to the analyzed taxocene, or of management irrelevance, by excluding those rarities that might have indicated either invasion or extirpation.

---

[38] Greenstreet and Rogers' (2006) observed decline was, in any case, largely between the catches of surveys in the 1920s and those conducted decades later (Greenstreet & Rogers 2006: Fig. 5). There must necessarily be doubts about the consistency of protocols and selectivities across that period.

In an analysis of data from the first three decades of Scotian Shelf groundfish surveys, Shackell and Frank (2003) binned the data into 300 km^2 squares (giving two or more sets per square) and then applied a novel alternative to rarefaction. They fitted a log-linear model to the relationship between the count of recorded species and the number of survey sets in each bin (though they scaled survey effort as 0.0404 km^2 swept-area per set). Residuals from the model fit were taken as indicating higher and lower richness than the norm for the survey area. While that avoided the data discarding of rarefaction, it could not escape the effects of varying steepnesses of Taxon Sampling Curves, made all the more serious as some of the 300 km^2 squares had seen no more than two survey sets in the 31 years of analyzed data, while none had seen more than 35 (Shackell & Frank 2003). Thus, the residuals may have been indicative of some form of small-scale α diversity but any relationship between them and the asymptotic *S* for the same squares must be very doubtful. Shackell and Frank's (2003) analysis indicated an unsurprising tendency for low counts of recorded species on the tops of the offshore banks and higher ones along slopes. It also found unexpectedly high counts in the Bay of Fundy and Sydney Bight but it remains unclear whether those areas are truly richer in fish species than most of the Shelf (i.e. their γ diversity at 300 km^2 scales is higher) or whether the higher counts resulted from fine-scale patchiness increasing sample species density (i.e. α diversity at 0.04 km^2 scales is higher).

In short, the curvilinearity in species-area relationships and the complexities of α, β and γ diversities preclude the direct use, in management applications, of sample values of species richness or species density (or indeed of ecological diversity *sensu stricto*) and most use of their means. Sample values must be used to derive community values, following the approach presented in Section 4.2.4, and only the latter used in generating management advice. Unfortunately, the derivation is not simple and requires far more data than will usually be available for any one small area. Attempts to evade those challenges will, however, lead to advice that is at best highly uncertain and often completely wrong.

4.2.6 Reporting Species Richness

One great advantage of richness is that its fundamentals are immediately understandable by non-technical audiences, yet the parallel disadvantage is that its complications and limitations are not recognized – indeed, are obscured by the apparent simplicity of the metric. Hence, whenever a value for richness is reported, it should be accompanied by sufficient caveats to prevent its misuse.

In research work, it is sometimes appropriate to compare richnesses across communities which share few, if any, species in common. In biodiversity mapping and monitoring, however, there will almost always be broad overlap of the compositions of the communities compared. Most ecologists will then respond to observed variations in richness by looking at species lists and finding which species are present in one community but absent from another. In many settings, it would therefore be well to accompany any report of differing richnesses with lists of species lost and gained – though it would be well to also report gains and losses when richness is unchanged. That, however, would be to move away from the strictures and limitations of ecological diversity, replacing them with a much more meaningful examination of species lists.

4.3 ECOLOGICAL DIVERSITY *SENSU STRICTO*

Many indices of ecological diversity *sensu stricto* have been suggested over the past half century. Their number has perhaps been exceeded by that of reviews of them. While some of those metrics continue to have value for particular research purposes, Jost's (2006) return to the Hill Numbers has successfully cut through the complexities where routine trend monitoring and spatial mapping in support of management are concerned. Jost's (2006) ${}^{q}D$ uniquely provides quantitative values for ecological diversity that behave as the intuitive expectations of non-specialists do, eliminating the massive misinterpretations that are all too easy with other metrics.

Within the range of ${}^{q}D$, the Inverse Simpson index (${}^{2}D$) has much to commend it, including that it can be estimated with less sampling effort than is required when $q < 2$[39]. However, that reduced effort is a result of the index being relatively

---

[39] Rosenzweig (1995), who preferred to define "diversity" as what others call richness, only saw value in the indices of ecological diversity *sensu stricto* as means for obtaining quantitative representations of communities from very small sample sizes (although, given larger samples, he did not admit the need to determine either asymptotic or rarefied S before comparing richnesses). He recognized the deficiency of all forms of the Simpson index, for his purpose, because of its dependence on N – which he termed a "bias". Dismissing what is otherwise universally known as the Simpson index of concentration as a "biased form", he then recommended a supposedly sample-size independent alternative:

$$\sum_i \frac{n_i^2 - n_i}{N^2 - N}$$

which he proceeded to treat as the "Simpson index". Unfortunately, that expression is not independent of N. Consider two data sets identical in every way except that one contains c times as many individuals as does the other:

insensitive to rarities, which are a significant focus of the policy concern over biodiversity. ${}^{0}D$, which is to say S, also has merit but that is recommended to be reported as richness and cannot usefully also stand as ecological diversity *sensu stricto*. Magurran (1988) argued for an index around ${}^{0.5}D$ as being the most sensitive, which would merit consideration if supported by further research but the metric would be unfamiliar and consequently harder to explain. Potentially multiple measures could be reported, such as ${}^{0.5}D$, ${}^{1}D$, ${}^{2}D$ and even ${}^{\infty}D$, the latter as a measure of dominance, rather than diversity. However, as has been argued repeatedly through this report, none of these metrics can carry much ecological information. To overwhelm end-users with a multiplicity of meaningless values would not be helpful.

Thus, we have suggested above (see Section 3.1.3) that the appropriate metric of ecological diversity *sensu stricto* for use in biodiversity monitoring and mapping is Jost's (2006) ${}^{1}D$, which is to say the exponential of the familiar Shannon index: $e^{H'}$, also called the Exponential Shannon index or EXP H'. Apart from the simple process of exponentiation, measuring ${}^{1}D$ comes down to determining a value for the Shannon index, H'. That has been the most widely used metric of ecological diversity *sensu stricto* and probably the second most widely employed of all measures of biodiversity, after species richness. It is familiar, as are the challenges in its quantification. Unfortunately, Shannon's has also been the most widely abused of the indices of ecological diversity *sensu stricto* – the abuse being of such long standing that Pielou (1966a, b) offered a corrective more than 45 years ago and presented the same points in her graduate and undergraduate textbooks a decade later (Pielou 1974, 1977). The misuse of Shannon's work in ecology has continued nonetheless. Indeed, as ecological diversity passed out of front-line research work and was left to routine operational tasks, such as environmental assessments, it was the Shannon metric that was "hailed by ecologists as a 'magic bullet'", as Washington (1984) put it. So bad has the situation become that correct applications of the index have been rather rare. There was some excuse in the 1960s and through to the mid-1970s, since the necessary calculations for correct application were tedious with the computing capabilities then available to biologists. Ten years later, those barriers had disappeared and yet Magurran (1988) was sufficiently unfamiliar with Pielou's points that the account of them in her influential textbook was

---

$$\sum_i \frac{(cn_i)^2 - cn_i}{(cN)^2 - cN} = \sum_i \frac{cn_i^2 - n_i}{cN^2 - N} \neq \sum_i \frac{n_i^2 - n_i}{N^2 - N}$$

Rosenzweig (1995) claimed that the independence of his index from sample size had been established by Nei and Roychoudhury (1974) but I cannot find that those authors mentioned Rosenzweig's (1995) metric, nor any that can be converted into it.

For determining diversity from small sample sizes, Rosenzweig (1995) also recommended the α parameter of the logseries on the grounds that "Small, incomplete samples of other distributions almost fit the log-series". That small samples do not provide enough data to demonstrate significant departures from some model is not, however, a justification for relying on the extrapolation of that model.

thoroughly confused, perhaps contributing to the on-going misunderstanding of Shannon's metric.

4.3.1 Determining Sample Diversity

The Shannon index has its roots in a measure of entropy (or information content), developed by Claude Shannon in the 1940s for use in information theory and subsequently adopted or re-invented for application in other sciences, not just ecology[40]. It is simply:

$$H' = -\sum p_i \log_b p_i$$

where p_i is the relative abundance of the i^{th} of S species in the taxocene of interest (after being filtered through the survey protocol) within the bounds of the community under consideration. "Abundance" is often taken literally to mean numerical abundance but biomass or other units can be used, as explained above. The logarithm can be to any base and hence the value of b must be reported whenever values of H' are given. In the past, b was typically set at 2, e or 10. If the exponential of H' is to be ^1D, however, b must equal e and hence that should be the standard in all future work.

Shannon's focus in information theory was on the entropy of a code, not that of a single message, and his index is defined only for very large sample sizes – strictly infinite numbers but certainly large enough that the abundance of the rarest species does not hover between zero and one individual. In ecological applications, that makes it a useful measure to apply to a community but not to a single sample, from which most rarities will (almost inevitably) be missing (though Pielou (1975) warned that the analogy between codes and ecological use of the same equation is weak and sought alternative arguments). The appropriate entropy measure for a sample is the Brillouin index:

$$H = \frac{1}{N}\log_b \frac{N!}{n_1!n_2!\ldots n_S!}$$

where n_i is the number of members of the i^{th} species in a sample or collection and N is the total number of individuals summed across the S species in that sample or collection (Pielou 1966b, 1974, 1975, 1977; Maurer & McGill 2011). Brillouin's index converges on Shannon's as N approaches infinity (Pielou 1975; Maurer & McGill 2011). Being determined without sampling error, as the collection in hand is itself complete and not a sample from something larger, values of sample H have no variance (Pielou 1966b, 1974, 1975, 1977).

---

[40] The index is often called "Shannon-Wiener", "Shannon-Weiner" (a mere spelling mistake) or "Shannon-Weaver". Spellerberg and Fedori (2003) explored the complex publication history of Shannon's work, showing that Warren Weaver had no connection with the development of the index (although both of their names were on the booklet which first presented it to a wide audience). Shannon himself acknowledged his debt to Norbert Wiener but did not share authorship. It is becoming normal to refer to the index using Shannon's name alone.

In the era before abundant computing power, calculation of the many factorials in Brillouin's equation discouraged its proper application and ecologists turned instead to the approximation:

$$H' \approx H'' = -\sum {}^{n_i}\!/_{N} \log_b {}^{n_i}\!/_{N}$$

which is a (mildly) biased estimator of *H'*. As a first approximation (neglecting higher-order terms) and provided that *H"* was calculated using natural logarithms, the bias can be corrected by adding (*S*-1) / 2*N*, where *S* is the (generally unknown) number of species in the community, not that in the sample (Pielou 1966a)[41]. The correction also requires an assumption that $N > S$ (Chao & Shen 2003), which is generally true for most data sets. That restriction limits the corrector to a maximum value of < 0.5[42]. Chao and Shen (2003) have offered a considerably more complex alternative estimator, along with one for the elusive *S*, but it is not clear that those represent a sufficient improvement in performance as to justify their use.

Unfortunately, avoidance of factorials was still emphasized by Magurran (1988), years after they ceased to pose any computational problem, while as late as 2004 she described Brillouin's Index as "more time consuming to calculate" (Magurran 2004), which was absurd given then-available computing power. Perhaps in consequence, to this day *H"* is routinely used in place of *H* – some three decades after the objections to Brillouin's metric became irrelevant. Indeed, *H"* continues to be widely misrepresented as being Shannon's metric, not just an approximation to it.

While *H"* has likely been used as a substitute for *H* and *H'* out of ignorance far more often than it has by considered choice, Peet (1974) drew a fine distinction between the proper way to measure entropy in a finite sample, which he agreed is Brillouin's index, and the best measure of ecological diversity when that is represented by entropy measures but not defined by them. He noted that, when used as a measure of sample diversity, Brillouin's has the disadvantage that its value increases with sample size. That is the expected behaviour of entropy, the sample value of the Brillouin index increasing (non-linearly) towards the corresponding value of the Shannon index, as sample size increases towards infinity. It is, however, an undesirable property in a measure of ecological diversity *sensu stricto*. Peet (1974) therefore preferred *H"*. Yet the diversity of a small sample is genuinely less than that of a larger sample drawn from the same community, the difference being correctly reflected in the sample-size dependence of *H*, and if the diversity of a sample (or even the mean of multiple sample diversities) is needed, the Brillouin index is the appropriate one. Thus, what Peet (1974) appears to have sought was not sample diversity but something larger, of which the sample diversity was but an indication. If so, he should have estimated community *H'*, using the approaches outlined in the next section.

---

[41] Buckland *et al.* (2011b) gave the correction as subtraction of (*S*-1) / *N* and implied that it is a corrective for bias in *H'* (rather than *H"*). Those suggestions were erroneous. Magurran (2004) provided the appropriate correction, including the first two of the higher-order terms, but showed them as the modifiers of *H'* needed to achieve *H"*, though she labelled the result as *H'*.

[42] Failure to abide by that restriction can result in overestimation of *H'* – which Chao and Shen (2003) achieved when they drew a sample of 50 individuals from simulated populations containing 100 species. Samples of 75 and 100 individuals did not always produce overestimates (depending on other characteristics of the simulations) but were nevertheless misleading representations of the bias corrector. Hence, Chao and Shen's (2003) conclusions concerning correction of the bias in *H"* must be set aside.

The use of factorials in Brillouin's index does confine calculation of *H* to integer data and hence encourages use of numerical abundance measures. It can, nevertheless, be applied to data on biomasses or other measures if those are rounded to some standard level, such as the nearest gram, before analysis and that level treated as the unit for counting "abundance".

4.3.2 Determining Community Diversity

When working with species richness, most ecologists are familiar with the nature and purpose of a species accumulation curve. Few would mistake the richness of a sample, or even the pooled richness of all of the samples in a database, with the asymptotic *S* of the community from which the samples were taken. Yet, values of the ecological diversity *sensu stricto* of a single sample or, more often, that of pooled samples are routinely (or "traditionally": Jost *et al.* 2011) reported as though they were community values. As Maurer and McGill (2011) have recently stressed: "this is incorrect". As is the case for sample richness, a value for sample diversity is only a single datum of limited interest, a very local α diversity, while the requirement is almost always for some community γ diversity (cf. Pielou 1966b, 1974).

For ${}^{q}D$ with $q \geq 2$, the effects of rarities on measured diversity are relatively small and hence community diversity can be adequately estimated from moderate numbers of samples. Indeed, nearly-unbiased estimators are available for that purpose (Jost *et al.* 2011). In contrast, ${}^{1}D$, the metric of ecological diversity *sensu stricto* recommended for use in biodiversity monitoring, is desirable for that purpose in part because it is sensitive to the presence of rarities. Those necessitate larger numbers of samples and complicate quantification when the index is applied to a community – as they also do with measurement of richness. Hence, to obtain community *H'* from sample *H*, one must proceed through a process analogous to the preparation of a Taxon Sampling Curve and its extrapolation to an asymptote.

Pielou (1966b, 1974, 1975, 1977) was the first to suggest a solution – one now known as "Pielou's pooled quadrat method"[43], which is closely analogous to

---

[43] Unfortunately, Pielou was probably inadvertently responsible for much of the confusion which has followed. She wrote of both "random" and "representative" sampling without always being clear what she meant. In one of her early papers (Pielou 1966b), she offered a method for estimation of the value of the Shannon index for a large sample (such as a plankton sample),

(Gotelli & Colwell (2001) said "effectively the same" as) the preparation of a sample-based species-accumulation curve[44]. She began by randomly ordering all of the samples and then created a series of growing data pools, the first pool containing the first sample alone, the second containing the first two samples, the third the first three and onwards until the last pool contained all of the available samples. Brillouin's *H* was calculated for each pool and the result plotted against the number of samples accumulated in that pool, generating a curve that increases towards an approximate asymptote, though with considerable variability. The asymptotic behaviour arises because the addition of further samples both reduces the observed relative abundances of those rare species which chance to be taken in early samples and adds further rarities. After sufficient samples have been collected, few further species will be added but the relative abundances of those taken early will continue to fall towards their true p_i, thus bringing pooled *H* towards a stable value. In contrast to a species accumulation curve, which nudges upwards when a sample adds another

---

when that is randomly subsampled in the laboratory and only the specimens in the subsamples fully identified and counted. Pielou (1966b) did not suggest that that method could be used for samples gathered in the field, where selection of specimens can very rarely be randomized (even if the distribution of sampling effort can be). Subsequent authors have, however, too readily described Pielou's (1966b) estimator for use with laboratory subsampling as one suitable for estimating community *H'* from randomly-collected field samples. Perhaps in consequence, they have ignored the very different approach that Pielou (1966b) did recommend for that kind of data.

Some authors (e.g. Magurran 1988, 2004) have further misunderstood this issue of randomness and have proclaimed that the Brillouin index is appropriate when the survey gear takes a biased sample of the organisms in its path, in the sense that a trawl net has different selectivities for different species, whereas *H"* should be used when the data come from some survey protocol that returns relative abundances equal to those in the community. No index of a dataset can ever be capable of blindly correcting for such structural biases in data to which it is applied. In reality, both Brillouin's *H* and *H"* are just as applicable to data gathered from the catches of selective gears and survey protocols (almost the only kind of gears and protocols that exist) as they are to hypothetical data from a non-selective survey.

Yet other authors have confused the bias in using *H"* as an estimator of *H'* with the bias arising from failure to complete enough sampling to obtain an asymptotic value of community *H'*. Even Chao and Shen (2003) presented the standard correction for the former as though it is a solution to the latter. In reality, the two biases are entirely different in both cause and magnitude.

Magurran (1988) did understand the issues at hand sufficiently to identify the difference between pooled-sample diversity and community diversity as a "more substantial source of error" than the use of *H"* as an approximation for *H'* but she did not offer any solutions for the former – not even the one that Pielou (1966b) had provided two decades before. Instead, Magurran (1988) presented Pielou's pooled quadrat method only as a way of determining required sample sizes and (mistakenly) as a means of coping with non-random sampling. Her presentation of those issues had not improved fifteen years later (Magurran 2004).

[44] Maurer and McGill (2011) noted that estimation of community diversity necessarily requires examination of the behaviour of the chosen metric of diversity among many samples from the community, though they did not explicitly link that with Pielou's pooled quadrats method. They did insist that such examinations involve an assumption that there is no temporal change in the community during the sampling period, which they considered to be problematic as most ecological units are open. The assumption is required but no more so than when preparing a Taxon Sampling Curve for the estimation of either asymptotic or rarefied *S*. Indeed, the use of any dataset for determination of diversity values implies an assumption that the values describe diversity as it was during the period of data collection. That should be addressed through the definition of the temporal bounds of the community under study.

species but otherwise runs horizontal, Pielou's curves decline between the upward nudges. Hence, when sufficient samples have been pooled, they show irregular variations around a horizontal line, instead of a true asymptote.

The value of *H* corresponding to that imaginary horizontal line, effectively the asymptote of the curve, is also the value that pooled *H* would take if very large numbers of samples had been gathered and were included in the pool. Since *H* converges on Shannon's *H'* as *N* approaches infinity, the asymptote is an unbiased estimator of *H'*.

With the limited computational capabilities of the time, Pielou (1966b) was confined to judging, by eye, the minimum number of samples needed for the pseudo-asymptote to be achieved and then averaging the *H* values for all pools containing numbers of samples greater than that minimum. The result of the averaging, $\overline{H}$, was taken to be the estimate of Shannon's *H'*, with sampling variance:

$$\mathrm{var}\left(\overline{H}'\right) = \frac{1}{\nu(\nu-1)}\left(\sum H_k{}^2 - \nu\overline{H}^2\right)$$

where ν is the total number of samples and H_k the value of Brillouin's index for the pool of the first *k* samples (Pielou 1966b, 1974, 1975, 1977).

When it has not been ignored or misunderstood, Pielou's approach has been improved but not, in its essentials, replaced. One obvious deficiency with the original form is that its results can be sensitive to the random ordering of the samples. An early solution was to work through the procedure several times, starting from alternative random orders of samples, the median or mean of the various $\overline{H}$ being taken as the estimate of *H'* (Heltshe & Forrester 1983). A second weakness is Pielou's subjective manner for determining a minimum number of samples needed for *H* of the pooled data to estimate asymptotic *H*. With modern computational power it ought to be straightforward to calculate *H* for pools of all possible pairs of samples, all possible trios and so on, followed by determining the asymptote statistically instead of by eye – much as has long been done with Taxon Sampling Curves. It seems that nobody has taken up the challenge of developing such a variant, perhaps because there has been no interest in "rarefying" *H'*.

In any case, the need for those limited improvements was deleted by the solution to a third deficiency in Pielou's original approach. In her method, once it is determined that at least κ samples are needed to reach the pseudo-asymptote, $\overline{H}$ is taken to be the mean of the values of *H* for each of ($K-\kappa+1$) pools of samples, where K is the total number of samples in the available dataset. That uses the available data inefficiently, since the ($K-\kappa+1$) values of *H* that are averaged are based on only κ, $\kappa+1$, $\kappa+2$ … K samples respectively. Zahl (1977) suggested jackknifing the data instead, thus creating K pools, each of $K-1$ samples. Provided that $K-1 \geq \kappa$:

$$\overline{H}' = \frac{1}{\mathrm{K}} \sum H_i$$

where H_i is the value of the Brillouin index for the i^{th} of the K pools[45]. As expected for a jackknife estimate, the variance of $\overline{H}'$ is lower than with Pielou's method:

$$\mathrm{var}(\overline{H}') = \frac{1}{\mathrm{K}(\mathrm{K}-1)} \sum \left(H_i - \overline{H}\right)^2$$

(Heltshe & Forrester 1983). The estimates of $\overline{H}'$ are little different: If the numbers of individuals in each sample were equal, each of the K jackknife replicates would yield the same value of H as would the asymptotic value for one of the possible random-ordering alternatives of the Pielou methods (Heltshe & Forrester 1983). The jackknife procedure is, however, considerably simpler to apply and uses the data more efficiently – provided that it is certain that $\mathrm{K}-1 \geq \kappa$.

Since κ cannot be known *a priori*, however, it is still necessary to draw up Pielou-type plots of H against the number of pooled samples, using multiple random sample orders. Fortunately, exact determination of κ is not required, provided that $\mathrm{K} >> \kappa$, and hence even subjective judgements of the asymptote are fully sufficient. If inspection of the curves shows that $\mathrm{K} < \kappa$, either the data must be grouped into larger units (e.g. longer time periods or larger map areas), more data must be gathered or else the attempt to estimate community diversity abandoned.

Once community H' and confidence intervals around it have been estimated, the values can simply be exponentiated to achieve ${}^1D$, the required measure of ecological diversity *sensu stricto*.

4.4 EVENNESS

In its original conception, evenness was supposedly co-equal with richness as a summary of a Species Abundance Distribution and as a foundation of ecological diversity *sensu stricto*, which was the summary of both richness and evenness into a single scalar value. In practical applications, moreover, evenness is found to be at least as useful a marker of differences among communities, or of anthropogenic disturbance of a community, as are richness and ecological diversity *sensu stricto* (e.g. Dornelas *et al.* 2011; Greenstreet *et al.* 2012a). Yet evenness is often treated as an afterthought, when it is not ignored altogether, being much less frequently reported or considered in environmental assessments or field reports than are the other two measures.

---

[45] Magurran (1988) presented Zahl's (1977) jackknife method as a preferable alternative to Pielou's Pooled Quadrat Method but failed to mention the requirement that $\mathrm{K}-1 \geq \kappa$. Without that, the jackknife approach risks severely underestimating H'. With it, Zahl (1977) provided an alternative for only one step in Pielou's process, not for the whole.

More surprisingly, first Heltshe and Forrester (1985) and much later Chao and Shen (2003) failed to note the requirement. Each pair of authors applied the jackknife to assorted sample sizes without ensuring that $\mathrm{K}-1 \geq \kappa$. Hence, their claims of bias in the jackknife estimates must be dismissed as without foundation.

In the early work of the quantitative ecologists, that was perhaps understandable: ecological diversity *sensu stricto* was a better summary of a SAD than, and hence a replacement for, richness, while (at a conceptual level) evenness was merely the difference between those two – "a comparison between the diversity as measured by the total number of species, and the diversity as measured by some other statistic", as Hill (1973) put it. Once ecological diversity *sensu stricto* could be quantified, there was little reason to emphasize either of its component parts and thus evenness could be set aside. As understanding of these measures has broadened and the need to interpret changes in ecological diversity *sensu stricto* in terms of its component parts has been understood, evenness should have been given equal status to richness but often has not been – though Magurran (2004) did give them matching prominence in her textbook. Notably, where richness and ecological diversity *sensu stricto* are quantified across a theoretical range of zero to infinity, indices of evenness are usually designed with a range from zero to one, as though evenness were a modifier of richness, rather than something which stands in its own right. Indeed, while there have been attempts to develop indices of evenness that stand alone, none have proven to have very satisfactory behaviours (Maurer & McGill 2011). The useful metrics of evenness are all ways to scale some measured value of ecological diversity, removing the effects on it of richness.

The earlier preference was to quantify evenness by scaling ecological diversity *sensu stricto* by the maximum that the chosen index could take in a hypothetical community with the same richness (i.e. the diversity of a community of S species in which each has the same relative abundance, $1/S$). Such metrics take the form D / D_{max}, where D is the measured value of diversity. That approach was still recommended in the mid-1970s by Pielou (1974, 1975), particularly in her evenness measures J', for use with the Shannon index, and J, for use with Brillouin's H. More recently, the D / D_{max} approach has been promoted by Magurran (1988, 2004). Hurlbert (1971), however, objected to evenness metrics of that form because the minimum possible value of many indices of ecological diversity *sensu stricto*, given a community of S species, is not zero. He argued that evenness indices should take the form $(D - D_{min}) / (D_{max} - D_{min})$, where D_{min} is the value that the chosen diversity index would take if applied to a community of one abundant species and S–1 that are vanishingly rare.

For a fully censused collection, containing N specimens, with its ecological diversity *sensu stricto* measured using the Brillouin index, H, there is a further complication in that the ratios involved in the calculation of H_{max} may not yield integer answers, which would prevent the calculation of factorials. Thus, Hurlbert's (1971) measure of evenness corresponding to Brillouin's H was:

$$V = \frac{H - H_{\min}}{H_{\max} - H_{\min}}$$

where:

$$H_{\max} = \frac{1}{N}\log_b \frac{N!}{(X!)^{S-r}(Y!)^r}$$
$$H_{\min} = \frac{1}{N}\log_b \frac{N!}{(N-S+1)!}$$

X being the integer part of N/S, r/S the residual from that calculation and $Y = X + 1$. Since it is based on a complete census, V has no variance (Pielou 1977).

The minimum value of Shannon's H' for a large community (the only kind to which that index should be applied) is zero[46] and hence the equivalent measure of evenness is unchanged from its earlier, D / D_{max}, version, *viz*.:

$$J' = \frac{H'}{\log_b S}$$

with variance:

$$\mathrm{var}(J') = \frac{\mathrm{var}(H')}{(\log_b S)^2}$$

where S is the total number of species in the community (Pielou 1974, 1975, 1977). In the interests of clarity, J' should always be reported using the same logarithmic base as is used for the corresponding value of H'. As considered above, in many cases community S is unknown and hence evenness cannot be estimated. Too often, the number of species recorded in the samples has been substituted (e.g. Magurran 1988), which yields only an upper bound on evenness (Pielou 1977).

While H' and J' have their merits, for present purposes the preferred measure of ecological diversity *sensu stricto* is the Exponential Shannon index. For that, Heip (1974) proposed the evenness estimator:

$$E' = \frac{e^{H'} - 1}{S - 1}$$

which is simply the exponential equivalent to J', if that is interpreted as a $(D\text{-}D_{min}) / (D_{max}\text{–}D_{min})$ metric, since the minimum value of the diversity index, e^0, is unity – though Heip (1974) neither cited Hurlbert (1971) nor invoked H_{min} to explain his subtractions, apparently deriving them empirically. As a metric of evenness should, E' has a minimum of zero and a maximum of one, regardless of the value of S[47]. We have been unable to find a published variance estimator for E'. Since much of its uncertainty is likely to lie in the value chosen for S, determining realistic confidence limits would likely prove challenging.

---

[46] As a single species approaches being 100% of a community, its p_i approaches unity, while those of all other species approach zero. Since the logarithm of one is zero, $p_i \log p_i$ approaches zero for all of the species present and H' thus also approaches zero.

Beisel *et al.* (2003) objected to J' because, they claimed, it does not have the range zero to one. They did not justify that position and hence it is not possible to determine, with certainty, where their error lay.

Through the past four decades, there has been far less discussion of evenness and its metrics of evenness than there has of corresponding questions about richness and ecological diversity *sensu stricto*. Such debate as has occurred has not led to any practical advance over Hurlbert's (1971) and Heip's (1974) conclusions. Some examination of the issues raised is nevertheless appropriate here. Of particular note, in an influential contribution[48], Smith and Wilson (1996) tested fourteen proposed measures against what they regarded as four essential requirements and ten desirable features for an evenness index. The fourteen included Pielou's J' metric for Shannon diversity, Heip's for Exponential Shannon, three related to the Simpson index (Gini-Simpson, Inverse Simpson and logarithmic Simpson forms), three others that make less-direct use of indices of ecological diversity *sensu stricto*, plus five that were independent of such diversity metrics, the latter including two proposed by Smith and Wilson (1996) themselves. Unfortunately, those authors' conception of "evenness" differed in important ways from what has been generally accepted amongst ecologists (cf. Gosselin 2001) and some of their tests thus returned misleading results.

Smith and Wilson's (1996) first and most important requirement was that any evenness metric must be independent of richness, which is certainly demanded by normal understanding of the ecological concept of evenness. They chose to test that requirement by applying the fourteen indices to three sets of dummy "abundances" (e.g. 1479, 1, 1, 1), each of which sets was replicated anything from one to 40 times to create a "community" to which each index was then applied. Smith and Wilson (1996) supposed that a metric independent of S would yield the same value, regardless of the number of replications. Thus, in place of a trial of independence, they substituted a test of what Jost (2010a) has since termed "replication invariance". Smith and Wilson (1996) were neither the first nor the last to make that mistake (e.g. Hill 1973; Heip 1974; Gosselin 2001; Ricotta *et al.* 2001; Ricotta 2004; Tuomisto 2012). That it was a mistake is easily illustrated: A single replication of Smith and Wilson's (1996) first set of "abundances" represents a "community" of four "species" – one with a relative abundance of 0.9980 and the other three with 0.0007 each. That is very close to maximal unevenness with $S = 4$. With five replicates, however, there would be 20 species, five with relative abundances of 0.1996 and fifteen having 0.0001, which is not maximally uneven for $S = 20$. (Maximal unevenness would approximate to nineteen species with relative abundances of 0.0001 and the 20th with 0.9980.) In the extreme, as Jost (2010a) has noted, one could consider an example of just two species, one overwhelmingly abundant and the other vanishingly rare. Such a "community" would show the minimum evenness achievable with $S = 2$ and any useful metric of evenness should yield a score of zero, or close to it, when presented with those values. Should that "community" be replicated four times, however, it would produce a new one comprised of four species, each with relative abundance 0.25, and four others vanishingly rare. That would be an example of intermediate evenness which, in another of Smith and Wilson's (1996) tests, would be expected to yield scores from an ideal evenness metric of 0.5, not zero. Replication invariance, in short, is not synonymous with independence of measured evenness from richness and the five metrics which Smith and Wilson (1996) found to "pass" their test (including the two of their own devising) are actually not fully independent of *S*, though their dependence on it is so scaled as to balance the increase in evenness which comes from the replication. Whether any of the other nine metrics that Smith and Wilson (1996) examined is truly independent of *S* remains an open question, pending application of alternative tests.

---

[47] $(e^{H}-1) / (S-1)$ has a minimum value of $(1-1) / (S-1) = 0$. It has a maximum value of $(S-1) / (S-1) = 1$.
Magurran (2004) made an issue of some earlier work which had replaced Heip's (1974) E' with the deficient $e^{H'}/S$. With evenness thus defined, H' is the sum of the logarithms of evenness and richness. While Magurran (2004) appears to have seen deep insight in that relationship, the evenness is question is calculated merely as ${}^{1}D / S$. Given that, ${}^{1}D$ must inevitably be the product of evenness and S. That simple inversion does not carry enough meaning to justify using a defective measure of evenness.

[48] Dornelas *et al.* (2011), for example, followed Smith and Wilson (1996) without comment on the earlier authors' errors. Gosselin (2001) provided citations of other authors who had relied on Smith and Wilson (1996).

Furthermore, Smith and Wilson (1996) sought "symmetry" in the performance of the evenness estimators, such that a community with one abundant species and many rare ones would score the same as a community with one rare species and many abundant ones. Yet, the definition of minimum evenness for a community containing *S* species has long invoked a single abundant species surrounded by *S*-1 rarities, whereas a community with multiple equally-abundant species and only a single rarity is very close to being one with all species equally abundant – which would match the definition of maximal evenness. By way of example: a sample in which there are 1,000 individuals of one species and one each of 100 others would be considered highly uneven. However, a sample of 1,100 specimens in which 100 species each contributed 10 or 11 specimens and a single rarity was represented by a lone individual would be considered very even indeed. Hence, the three metrics which Smith and Wilson (1996) found to "pass" their test by showing "symmetric" behaviour (their own two plus the "NHC" index) must be rejected out of hand as grossly misrepresenting evenness (cf. Gosselin 2001; Beisel *et al.* 2003).

Eight of the fourteen metrics which Smith and Wilson (1996) examined successfully failed both of those inappropriate tests but two of them also failed a better-designed test for Smith and Wilson's (1996) second requirement: that reducing the relative abundance of the rarest species should reduce the index value.

The residual half-dozen indices met each of Smith and Wilson's (1996) remaining requirements and showed all the other features that they considered desirable, except that one showed a discontinuity in the slope of its response at maximum evenness, while all six failed to generate results around 0.5 when presented with data that Smith and Wilson (1996) thought represented intermediate evenness (e.g. $S / 2$ species equally abundant and another $S / 2$ vanishingly rare). The behaviour of evenness indices when confronted with such data has been an issue of prolonged debate, with Pielou's J' deprecated because it does not produce a value of 0.5 unless $S = 4$. Jost (2010a) has, however, shown that ln J' does perform according to expectations. He has also pointed out that the idea that evenness should be 0.5 when half the species are equally abundant and half rare cannot be extended to the case of $S = 2$, since one abundant species and the rest (in this case: one species) being rare is the definition of minimal evenness and should be scored as zero – as noted above. Jost (2010a) recommended that any change in what he termed "relative evenness" (corresponding to "evenness" as that has been understood since the work of Hurlbert 1971) should be visualized in terms of "steps", each involving an exchange of $N / 2$ individuals from rare to abundant species or *vice versa*. Starting with $S = 4$ equally abundant species ($J' = 1$), a single such "step" produces two abundant and two rare species, a second "step" leading to one abundant species and three rarities ($J' = 0$). Each "step" thus reduces J' by 0.5 and the intermediate community has $J' = 0.5$. Should $S = 8$, however, it would take three "steps" to progress from maximal to minimal evenness but only one to reach the intermediate condition of four equally abundant species and four vanishingly rare. Hence, the correct value for evenness of an intermediate community of $S = 8$ (1/3 removed from perfect evenness) should be 0.667 – which is the value of J' in that case (Jost 2010a).

Thus, there were five plausible indices of evenness amongst those considered by Smith and Wilson (1996): Pielou's J', Heip's index for use with the Exponential Shannon diversity index, indices for the Gini-Simpson and logarithmic Simpson indices, and Pielou's eveness metric derived from McIntosh's diversity index. While it cannot be said that those five have been thoroughly tested, at least Smith and Wilson's (1996) work gives no reason to reject any of them – those authors' conclusions to the contrary notwithstanding.

Each of those five metrics is linked to, indeed derived from, a particular index of ecological diversity *sensu stricto*. It would be possible to measure evenness by a metric linked to a different diversity index from that used in the same program but there is no obvious reason to do so. Indeed, deriving evenness from diversity largely reduces the former into a way of explaining trends in ecological diversity *sensu stricto* which deviate from those in richness. That explicatory role is greatly facilitated when the linked indices are used together. Hence, the choice of the Exponential Shannon metric for quantifying ecological diversity *sensu stricto* all but requires that Heip's index be used for evenness.

Amongst the few more recent studies of the topic, Ricotta *et al.* (2001) conducted a Principal Components Analysis of supposed "evenness" values generated by seven metrics, when they were applied to a data set on terrestrial plants. Those authors made much of the analytical output distinguishing one index from the rest. The distinctive one was, however, ${}^{\infty}D / S$. Since ${}^{\infty}D$ is itself an evenness metric (as ${}^{0}D$ is a measure of richness), it is hardly surprising that its division by S generated a measure that deviated from evenness. The other six indices, which included J' and metrics of the form D / D_{max} derived from the Gini-Simpson and Inverse Simpson indices of ecological diversity *sensu stricto*, all performed rather similarly and Ricotta *et al.* (2001) were unable to offer much ground for selecting among them.

Building on an idea first proposed in the 1970s, Gosselin (2001) suggested that the concept of evenness should be founded on the mathematical "Lorenz partial order", not as a convenient quantitative representation but as a definition, from which postulate he was led to accept replication invariance as synonymous with the independence of evenness from richness, as noted above, despite its clear failing when $S = 2$. Gosselin (2001) rejected Smith and Wilson's (1996) insistence on symmetry – not because of conflicts with long-established understanding amongst quantitative ecologists but because their requirement conflicts with the Lorenz partial order. He identified the evenness metrics derived as ${}^{q}D / S$ (for all q) and as the Gini-Simpson diversity index divided by S to be "Lorenz-compatible", while ${}^{q}D / {}^{q'}D$ with $q' > 0$, Heip's (1974) index and a number of others are incompatible. However, the failure appears to lie not in those metrics but in Gosselin's (2001) insistence on founding "evenness" in a mathematical form that fails to follow ecological understanding.

Gosselin (2001) did usefully explore a conundrum in the meaning of "evenness": when one species of a perfectly even community declines in abundance, leaving the other species unaffected, evenness should also decline until the one species becomes vanishingly rare. However, when the last individual of that species is lost (something essentially undetectable in almost all field surveys), evenness should return to unity. That discontinuity is intuitively unsatisfactory and Jost (2010a) later sought what he termed "continuity", by which he meant that, as the rarest species in a community containing S species declines towards zero abundance, evenness should decline towards the value that it would have with S–1 species, after deletion of the rarity. Unfortunately, such continuity is inconsistent with evenness, as that concept is generally understood, and Jost (2010a) did not (perhaps could not) identify any metrics of evenness that showed his "continuity". Intellectually uncomfortable though it may be, the existence of extremely rare, essentially undetectable, species does lower evenness and the available indices reflect that.

Ricotta (2004) restated much of Gosselin's (2001) argument but also extended the problems by noting that, if evenness is represented by the Lorenz partial order, a minimally even community (one with a single abundant species and S-1 that are vanishingly rare) necessarily becomes less even as S increases, calling into question the required independence of evenness and richness. Ricotta (2004) attempted to advance an alternative that might avoid both problems but they appear intrinsic to the concept of evenness and neither metrics which evade them, nor a definition based on the Lorenz partial order, are likely to be helpful.

Meanwhile, Beisel *et al.* (2003) examined much the same set of evenness measures as had Smith and Wilson (1996) but, besides avoiding the erroneous requirement for symmetry set by the earlier authors, they used real data on riverine benthos and simulations based on those data in their trials instead of fabricated "communities", though their means of judging the relative merits of the indices when applied to those data remain obscure. Curiously, Beisel *et al.* (2003) did not acknowledge the desirability of matching the chosen index of evenness to the one selected as a metric of ecological diversity *sensu stricto*. In the absence of that consideration, they recommended Pielou's J', Hurlbert's V, indices for the Gini-Simpson and logarithmic Simpson indices, and Pielou's metric of evenness derived from McIntosh's diversity index. It is unclear why they set aside Heip's index, which they found to be strongly sensitive to the presence of rarities, whereas none of their preferred alternatives showed more than weak sensitivity to the numbers of either rare, moderate or abundant species.

More recently, Jost (2010a) examined questions of evenness, as he has so much else within the scope of ecological diversity. Unfortunately, he made the mistake that he had previously accused others of (Jost 2006) by confusing the concept of evenness with the indices used to quantify it – a fault that Hoffman and Hoffman (2008) saw even in Jost's (2006) earlier work. Moreover, Jost (2010a) dismissed Hurlbert's (1971) point that evenness indices should take the form $(D\text{-}D_{min}) / (D_{max}–D_{min})$, choosing to denote those as measures of "relative evenness". In their place, he proposed that the metrics of evenness should be, in the symbols of his "true" diversity, ${}^qD / S$ which is to say ${}^qD / {}^0D$ – a return to the errors of the 1960s that Hurlbert (1971) had corrected long before. Having done so, Jost (2010a) proceeded to show that that his measures of evenness are not fully independent of richness, as the concept of evenness is required to be. More exactly, in the case of ${}^1D / S$ (the version appropriate to the Exponential Shannon diversity index), $e^{H'}$ has a minimum of one and hence ${}^1D / S$ has a lower bound of $1 / S$ – meaning that it cannot be fully independent of S, though it becomes less sensitive to that variable as richness increases. None of that was novel, however, having been evident to, and explained by, Heip (1974) long before. The same form of lower bound is seen with ${}^qD/S$ for other values of $q > 0$ (Jost 2010a).

Having established that degree of dependence, however, Jost (2010a) did not accept the long-known $(D\text{-}D_{min}) / (D_{max}–D_{min})$ solution but bizarrely declared that $S / {}^qD$ does not share the inverse of the same dependence seen in ${}^qD / S$. Persuading himself that he could thus eliminate the effect of richness on measures of evenness simply by inverting the latter, he recommended discarding conventional notions of evenness as something which varies from zero (minimal evenness) to one (perfect evenness) and replacing them with inverse ideas of "inequality", varying from one (perfect evenness) to infinity (minimal evenness). Continuing his confusion of metrics with concepts, Jost (2010a) then declared that ecological diversity *sensu stricto* cannot be decomposed into independent richness and evenness but richness (which is only a count of the species present) can be decomposed into independent diversity and evenness. If inversion could nullify dependence and if that dependence was anything more than a consequence of Jost (2010a) erecting a strawman, derived by ignoring four decades of progress in his topic, then such a wholesale revolution in ecologists' understanding of diversity might have to be considered. Since the one cannot and the other was not, Jost's (2010a) ungainly edifice can be discarded[49].

One outstanding problem with Heip's index of evenness, and with most others based on scaling of measures of ecological diversity *sensu stricto*, is that they depend on the value of (often unknowable) S. Indeed, Pielou's J' is inversely proportional to the input estimate of S and Heip's index is not far from that (being inversely proportional to S–1). When applied to depauperate communities, for which richness can readily be determined, that relationship poses no challenges. For very rich, open communities, it can make evenness impossible to quantify. That problem cannot be avoided by using a constant, dummy value for S. If so calculated, "evenness" would be directly proportional (or close to directly proportional) to ecological diversity *sensu stricto* and temporal or spatial variations in such a dummy "evenness" would merely reflect those in diversity.

Jost (2010a) considered that problem. He rightly rejected the common suggestion that measures of evenness should be restricted to fully censused plots (cf. Smith & Wilson 1996), since extrapolation to the community scale is usually needed. He investigated what he termed a "diversity profile" –a plot of ${}^qD$ against q– which forms a curve declining from S at $q = 0$, unless the community is perfectly even, in which case its "diversity profile" forms a horizontal line at S. The depression of the profile below the horizontal is thus a measure of evenness, while ${}^qD / {}^0D$ relates to the average steepness of the curve over the interval between zero and q. To avoid the need to know ${}^0D = S$, Jost (2010a) considered the option of calculating ${}^2D / {}^1D$ as a different index of the slope of a "diversity profile", an idea which Hill (1973) had suggested long before, but Jost (2010a) showed that it is an ambiguous indicator of evenness. The ambiguity could be avoided by using $\ln{}^2D / \ln{}^1D$ but Jost (2010a) stopped short of recommending that metric and did not fully explore its behaviour. Instead, he reasoned that, if a community were well characterized, with its ecological diversity *sensu stricto* firmly established, when additional survey work detected one or more extreme rarities, existing understanding of the ecology of the community would be, for all intents and purposes, unaffected, yet S would increase by the number of new rarities and measured evenness would decline approximately proportionately.

---

[49] Jost (2010a) examined the meaning of his $S/{}^qD$ measure of "inequality", leading to insights that are of some interest. It is only his attempt to replace evenness with "inequality" that should be discarded.

Jost (2010a) therefore suggested that, rather than examining the evenness of the entire chosen taxocene (as filtered through the selectivities of the survey protocol) in the community of interest, attention should be confined to a part of the taxocene / community unit that excludes extreme rarities and for which a truncated richness could be readily determined. He went so far as to lay out a semi-rational (but still partly arbitrary) means of doing so. That was an interesting proposal but not one that we can endorse. It was founded on an assumption that rarely-recorded species are ecologically unimportant, yet the scarcity of records can be as much a function of the selectivities of the survey protocol as they are of actual abundance, while even the latter is a poor indicator of the ecological significance of large, scarce top predators. Jost (2010a) further implicitly assumed that evenness of the ecologically significant members of the community would be an ecologically meaningful measure, which seems (at best) hard to substantiate. If evenness measures are merely descriptors of Species Abundance Distributions, which we would argue is the case, then the ability to generate repeatable, comparable and understandable values is more important than selective exclusion of ecologically insignificant species, while Jost's (2010a) means of selecting would retain a variable number of species, depending on the degree of anthropogenic depression of the abundant species. Besides, if the principal value of measures of evenness is to help in explaining observed variations in ecological diversity *sensu stricto* and if the latter is calculated using data on rarities, then any evenness metric that excluded those species would lose its point.

Finally, Tuomisto (2012) offered a contribution on evenness which, in fairness to its author, should be consigned to quiet obscurity. She sought to define (not merely quantify) evenness but did so from a starting point that "Deciding which measure to bestow the name 'evenness' on implies no opinion on which measure is most important or most useful, so the decision can be entirely based on convenience". From such a nihilist foundation, she chose to declare diversity to be nothing more complex than the product of richness and evenness, allowing her to define the latter as diversity divided by richness. Having swallowed selected parts of Jost's (2010a) argument, she then rendered evenness as ${}^{q}D / {}^{0}D$, despite acknowledging that that violated the required independence of evenness and richness. Aside from its algebraic simplicity, her alternative has no merits in its favour. There is no need to review here the other missteps in Tuomisto's (2012) paper.

We suggest that these various complications can be evaded by relying on pragmatism more than theory, no matter how intellectually dissatisfying that may be. Species richness is often very difficult, and sometimes impossible, to measure but it is tangible – simply the number of species in the taxocene and community of interest (as filtered through some survey protocol). Ecological diversity *sensu stricto* should be seen as a summary of a Species Abundance Distribution into a single scalar value, which can be expressed as a number of "effective species". Evenness is then the factor which scales the number of effective species to the total number of species, taking the form $(D-D_{min}) / (D_{max}-D_{min})$. It should be reported using the same measure of D as is used in reporting ecological diversity, *sensu stricto*. Since we recommend the use of the Exponential Shannon metric in the Department's monitoring and mapping of ecological diversity *sensu stricto*, evenness should be quantified using Heip's index.

4.5 EXPERIMENTAL DESIGN AND THE DRAWING OF INFERENCES

This Section 4 has emphasized the need for comparisons to be made among comparable values of ecological diversity – ones for the same taxocene, as recorded using the same survey protocol, within communities that have similar spatial and temporal extents, with the same metrics used for the quantification. All of that, however, is but to require that measurements are made in the same units, as length measurements might all be in metres, rather than some in millimetres and others in fathoms. At best, it is equivalent to requiring that all fish lengths be measured as fork lengths in centimetres, not a random mixture of fork, total and standard lengths.

If inferences are to be drawn from comparisons amongst such measurements of ecological diversity, however, the full complex of considerations that would be invoked when drawing inferences from any other kind of data must also be addressed – as Rodda (1993) emphasized with a tongue-in-cheek presentation of the sorts of errors which can arise when the need for appropriate experimental design is forgotten or, by his suggestion, deliberately set aside. There are basic statistical questions (distributions of error terms, statistical significance of any observed differences, power of tests etc.). Analyses of time series face issues of autocorrelations in time and of aliasing (cf. Dornelas *et al.* 2012). Mapping of spatial variations likewise faces geospatial considerations. Should the values of ecological diversity be used in experimental investigations of the effects of particular anthropogenic pressures, a more complex experiment would be needed, probably involving some form of replicated, randomized BACI design.

Those are all issues that go far beyond the scope of the present report, which is confined to the means of measuring diversity using indices. Ours is an important topic but so is the experimental design and that must not be forgotten when moving from theoretical treatments to practical applications, especially when it is necessary to infer the causality underlying observed changes (cf. Rice 2000).

5 MEASURING OTHER FACETS OF BIODIVERSITY

In Section 2, we argued that ecological diversity is measurable but is so because it is conceived (if not quite defined) as something that can be quantified by simple indices. The simplicity makes ecological diversity measurable, albeit not without considerable effort, but it also and necessarily limits the amount of meaning that the indices can carry. Biological diversity, in the sense of the CBD definition that is followed here, is very different. While it does not encompass all that is relevant to the conservation of ecosystems (as does environmental status), biodiversity is meaningful. It is not, however, measurable as a whole, though single facets may be. Yet they must be measured directly, as ecological diversity fails to serve as a convenient surrogate for any of them.

In this Section, we briefly examine approaches to that measurement which use indices, taking the principal data types or facets in turn. Unfortunately, and in contrast to ecological diversity, there is limited depth of experience with any of the metrics of other facets of biodiversity. Hence, it is not yet possible to provide the sort of detailed account of their correct use that is offered in Section 4 with respect to ecological diversity. We do offer more detailed comment on a few particular indices that have achieved high prominence in recent years. Our focus here is confined to the use of indices and, while we note the availability of more advanced alternatives, we make no attempt to review them. We thus face a question of whether the depth of meaning in the several facets of biodiversity can be captured by simple metrics or whether their effective measurement requires something more complex.

We do not stress the point in the following paragraphs but, for measurements of any facet to be comparable across time or space, they must be made on comparable units – something equivalent to the communities and taxocenes, filtered through specified survey protocols, required of measurements of ecological diversity. Meanwhile, if the comparisons are to have any ecological meaning, the indices used must have sufficient precision and sensitivity, while responding in a understandable way to changes in the measured ecosystem, and they must be applied through an appropriate experimental design. Establishing that a metric meets those standards requires a validation step (Borja *et al.* 2009), which in the past has too often been absent from index-development efforts. Although not repetitively stated through the rest of this Section, those requirements should not be forgotten.

5.1 SPECIES RELATIVE ABUNDANCES & COMMUNITY COMPOSITIONS

Data on the mix of species in a community can be reduced to the simplistic indices of ecological diversity that are considered above. However, those who set out to apply that approach often stray into more complex ways of addressing such data. Even the early quantitative ecologists soon veered towards examining Species Abundance Distributions directly, instead of summarizing them in one or more indices, an alternative that has been recommended through the decades (e.g. Magurran 2004). Hurlbert (1971) was one who did so, stressing that the structures of biotic communities cannot all be arranged in some uniquely-appropriate order on a single linear axis, such as might be represented by the value of one index. Distributions of the numbers of species within a community have many properties, any of which can be compared among communities, while the relative rankings of the communities will differ depending on the property considered. In a similar vein, Leinster and Cobbold (2012) advocated the use of diversity profiles (cf. Hill 1973; Jost 2010a) rather than single index values, since the former are considerably more informative. Their recommendation was for profiles of ${}^{q}D^{Z}(p)$, which is explained below, across all possible values of q but would apply equally to profiles of ${}^{q}D$. Unfortunately, while diversity profiles are useful as aids for experienced analysts in understanding differences in diversity across time or space, any type of profile is poorly suited to routine reporting. In a parallel departure from the simplicity of indices, Pielou (1975) recommended using species lists, with the species in descending order of abundance, whenever S is small, resorting to analysis of SADs only when the community of interest contains many rarities. Indeed, when they find variations in ecological diversity, many ecologists turn to examination of species lists, trends in the abundances of individual species and the autecology of the principal species, in order to interpret patterns in the summary metric (e.g. Magurran 1988; ICES 2011b). That calls into question the purpose of evaluating indices of ecological diversity at all.

Various alternative ways of examining species-abundance data have been developed (e.g. Dornelas *et al.* 2011), including dominance / diversity plots (i.e. species relative abundances, or their logarithms, plotted against species' rank-order or sometimes its logarithm), *k*-dominance plots (much as dominance / diversity plots but using cumulative species abundances), abundance / biomass curves ("ABC curves": as *k*-dominance plots but with cumulative biomass plotted on the same axes as cumulative abundance), and cumulative distribution function plots (cumulative percentage of species plotted against the logarithm of percentage of community total abundance). Those can be very useful in research work, and have been characterized as informative in capturing the effects of fishing on ecosystems (Rice 2000), *k*-dominance curves in particular being endorsed by Rice (2000) and ICES (2005). We recognize their value as analytical tools for use when assessing environmental status. However, the subjectivity inevitable in graphical interpretation and the absence of numerical outputs from the graphs renders them of dubious merit as measurements. There have been some attempts to quantify the comparisons of ABC curves, using the W statistic, and to extract statistics from other plots. Q, for example, is the inter-quartile gradient of a cumulative distribution function. There has not, however, been the depth of experience with those approaches to encourage confidence in their interpretation.

Indeed, the appropriate way to analyze data on species compositions and relative abundances is not some graphical device nor any oversimplified metric but rather through multivariate statistics, particularly the various kinds of ordination, which are capable of capturing far more of the information in a dataset. Several specialists in the measurement of ecological diversity have admitted the superiority of that alternative. McGill (2011), for example, stated that there "can be little doubt that sophisticated multivariate analyses are perhaps the most rigorous way of analyzing abundance data". He did not recommend them, however, preferring to emphasize the "long-standing popularity of simpler [...] approaches" – despite himself having shown the inadequacies of the latter. Setting aside the very extensive biological applications of multivariate statistics, in both ecological and taxonomic work, since they became widely available as packaged software *circa* 1980, McGill (2011) then bizarrely claimed that multivariate methods have been little used in the analysis of species abundances because of their complexity and the "training required" for their correct use! For their part, Dornelas *et al.* (2011) considered multivariate analyses to offer "potentially very powerful" means for detecting the effects of disturbance on biodiversity, which they do[50]. Similarly, Jost *et al.* (2011) started their contribution to a textbook on biodiversity measurement with an account of indices of similarity which, in direct contrast to those of ecological diversity, respond to commonalities and differences in the species compositions of two communities, instead of considering only relative abundances within each community, while remaining blind to species replacement, as the indices of ecological diversity do. Even Magurran (2011) recognized that, while temporal change in a community could be examined through a time series of diversity values, similarity indices would additionally capture species replacement. Being less focused on diversity metrics than those specialists are, over the past three decades the wider community of ecologists has simply applied suites of powerful multivariate techniques to species-abundance data. Even those who have seen fit to also calculate indices of ecological diversity have extracted much more information through their multivariate approaches, which have often been supplemented by examination of the data on individual species (e.g. Jennings *et al.* 1999; Callaway *et al.* 2002; Schratzberger *et al.* 2006, 2007; Rees *et al.* 2007). Nor is the suite of multivariate tools static – new and more powerful approaches especially suited to assessments of environmental quality, such as "principal response curves" (Pardal *et al.* 2004), continue to emerge.

In sum, once data on the relative abundances of species have been expensively collected, rather than merely calculating the values of indices of ecological diversity, much more efficient, informative and sensitive approaches, based on multivariate statistics, are available and should be used. Selecting the most appropriate options within that general approach lies beyond the scope of the present report but we would suggest that the central feature would likely be an ordination of the species abundances, followed by examination of temporal trends or spatial variations in mean scores on the first (or first few) orthogonal axes. That would be a far more productive course to follow than further pursuit of the sterile ideas of ecological diversity, which properly belonged to the pre-computer era.

---

[50] Dornelas *et al.* (2011) did raise a warning that analyses which treat differently named species as being different (as most multivariate approaches do, the aim being to utilize the information encapsulated in the names) cannot extend across ecosystems that have few or no species in common. That is a constraint on some research work. In mapping or monitoring, however, it is not an issue: if there had been substantial species replacement over time, that would be an important result even if the ecological diversity of the community remained unchanged.

5.2 COMMUNITY AND SPECIES ABUNDANCES AND BIOMASSES

Relative abundances contain information on the variability within a community defined in time and space, which is to say on a facet of biodiversity. Absolute abundances contain additional information on spatio-temporal variations that are often of great importance to ecological status but they do not add to the information on within-community variability. Thus, measures of abundance *per se* are not metrics of biodiversity in the sense of the CBD definition. One of the weaknesses of ecological diversity, however, is that (as it is generally understood) it ignores absolute abundance. Buckland *et al.* (2005) therefore sought new measures of "diversity" responsive to both overall absolute abundance and species' relative abundances, proposing a modification of the Shannon index which would have that property. That, however, meant wrapping richness, evenness and overall abundance into a single metric, which is unhelpful. Thus, Buckland *et al.* (2011a) later accepted that abundance and ecological diversity are better tracked separately, while McGill (2011) recommended reporting overall abundance (summed across species) alongside species richness and an evenness measure – an approach we endorse whenever ecological diversity is mapped or monitored across time. Similarly, Shin *et al.* (2010b) selected "total biomass of surveyed species" as one of eight "indicators" of ecological status of marine ecosystems.

Besides tracking overall abundance, as one measure of environmental status and a support to interpretation of trends in ecological diversity, the suite of indicators applied to a particular ecosystem should likely include various taxon-specific abundances, biomasses and related measures such as patch size and proximity (cf. Kenchington *et al.* 2012), including those of resource populations, species at risk and perhaps others besides – forage species and sensitive "indicator species" or "sentinel species", for example (but see Noss 1990; Rice 2003). In practice, the array of populations to be tracked, across time or space, will probably be determined more by the availability of data than by other considerations. Whether those abundances and biomasses fall within the scope of the CBD's biodiversity may be a moot point (though the abundance of habitat-structuring species can sometimes be taken as indicative of areas with higher ecological diversity: cf. Kenchington *et al.* 2012) but the need for monitoring of them is not. Indeed, one tenth of the "indicators" recently accepted for monitoring the CBD's Aichi Biodiversity Targets[51] concern either the abundance of selected species or the extinction risks that those abundances indicate. The few other indicators in that set which involve monitoring of biodiversity (as distinct from anthropogenic pressures, conservation measures and other related issues) concern the spatial extent of biomes, ecosystems and habitats.

---

[51] UNEP/CBD/COP/DEC/XI/3, December 2012.

Of interest here is whether those measurements of abundances and biomasses should be reported as a complex of many time series (or many maps, in the case of spatial variation) or else combined into a single series (or map) for each taxocene in the monitored community. More exactly, since both formats are likely to be used in any real-world application, the issue is where the emphasis should be placed when the time series or maps are presented to decision-makers.

Reporting of abundances and biomasses summed across a taxocene has not been without supporters and, in highly-diverse tropical and subtropical areas, there is merit in the idea. For Canadian waters, however, we strongly recommend playing down such summary measures. Since aquatic ecosystems tend to maintain productivity in the face of anthropogenic pressures (see Section 2.3), community biomasses tend to be insensitive, even stable, despite massive changes in population-specific abundances. That is not to say that time series of summed biomasses should never be presented. They should and will be, though they are best presented with the cumulative contributions of various species indicated. Nor should whole-taxocene measures be set aside if nothing better is available: it would clearly be better to report a temporal trend in overall bivalve biomass than to offer no report at all. However, where data on the abundances of individual species or populations are available, emphasis should be placed on them. That recommendation does not propose anything that would seem novel to DFO's fishery managers, who expect scientific advice not just on each species but on each "stock". Likewise, the Department's species-at-risk staff expect abundance information by population, not community. Those expectations should be carried over into monitoring and mapping of biodiversity.

Dulvy *et al.* (2006) suggested the reverse, a single index (which they termed an "indicator of marine biodiversity") summarizing the abundances of a selected subset of the species taken by a single survey series. Further, instead of tracking changes in abundances directly, they proposed scoring decreases in abundance since some chosen year early in the survey series, declines of ≥90% being scored 3, ≥70% 2 and ≥50% 1, all other species being scored zero. Their proposed index was a simple average of those scores across the selected species. As applied by Dulvy *et al.* (2006), once a species was given a positive score, it did not revert to zero until fully recovered to the abundance observed early in the time series. They avoided the need to provide negative scores for species that increased in abundance by excluding those from their selected subset. We see no discernable advantage in that approach, relative to straightforward tracking of abundances, and Dulvy *et al.* (2006) did not suggest any.

A rather more useful measure, the "Living Planet Index" ("LPI"), has been advanced by the BIP for both global and regional application and has been adopted by the CBD. It tracks change in the geometric mean of abundances of vertebrate animal species, without regard to species richness or other aspects of ecological diversity. Each time series is standardized to a value of unity at a selected start date, conventionally 1970. To allow the use of all available estimates of trends in species abundances, including both measures proportionate to absolute abundance and runs of data that commenced after the start year, the series is constructed by finding annual geometric means of year-over-year proportionate change in species' abundances, each such mean being applied to the previous-year's relative overall abundance. Time series are frequently displayed separately for each class of vertebrates (McRae *et al.* 2008).

The LPI can provide an overarching summary of available quantitative abundance information, which may have merit as a crude "scorecard" when communicating to lay audiences. Its deficiencies as a scientific measure are, however, severe. The non-random availability of abundance estimates (e.g. for fish: emphasizing resource species; for marine mammals: emphasizing species at risk), the use of partial runs of estimates (altering the species tracked across time), the equal effect on the Index of a 10% decline in either an ecologically-key species, an undistinguished species or an endangered species, and the arbitrary allocation of LPI = 1 to 1970 all constrain interpretation of trends. In combination, they render the LPI meaningless. Moreover, the reliance on geometric means prevents the use of any time series that shows zeros in some years. Chance fluctuations in observed numbers of rare species are given the same weight as proportionate shifts in abundant species, making output values unstable. Setting of the index for the baseline year to unity (with no variance), when real abundances for that year were only estimated, complicates the variance structure of the rest of the time series (Buckland *et al.* 2011a). In practical applications, there is also a serious risk of biasing outcome to match an analyst's expectations, through selection of which time series of abundances to include. The overall LPI for Uganda, for example, tracks the national trend in mammals, rather than a more optimistic one for birds, simply by being built on data for 26 mammal species but only nine birds. In contrast, the Canadian national LPI tracks that for birds, rather than the very different ones for mammals or fish. The Canadian fish LPI (McRae *et al.* 2008) shows the severe decline from the late 1980s into the new century that is all too familiar for Atlantic groundfish but which is unrepresentative of trends in many other species – suggesting a bias in available data. Therein lies the most serious deficiency in the Index: useful management responses must be targeted where particular declines are occurring and no measure accumulated across species and areas can identify those.

Indices that take the form of ratios of the abundances of different groups of organisms may be more useful. The ratio of planktivorous fish to ichthyovores, for example, could serve as a trophodynamic indicator (cf. Cury *et al.* 2005). Shin *et al.* (2010b) suggested the proportion of predatory fish (defined as all piscivorous fish species, plus those which eat invertebrates larger than 20 mm) as an "indicator" of biodiversity. Meanwhile, ratios of pelagic to demersal fish have been of research interest. However, it is rarely possible to track ratios or proportions of abundances or biomasses in a community, since any such measure is dependant on the availability of estimates from routine, standardized surveys and very rarely will the selectivities of the survey protocols be known for all of the species concerned. Ratios of relative abundances can show temporal change nonetheless but care is needed in interpretation of any observed trends. For example, it might be tempting to monitor a groundfish : herring (*Clupea harengus*) ratio in the catches of the Gulf of St. Lawrence or Scotian Shelf groundfish survey series. Should that ratio change, however, it would not be clear whether the shift was in the demersal : pelagic ratio or in the fraction of the pelagic biomass that was composed of herring, its principal competitors (mackerel, *Scomber scombrus*, and saury, *Scomberesox saurus*) having much lower selectivities in groundfish nets.

5.3 PHYLOGENETIC AND TAXONOMIC DIVERSITIES

Early students of ecological diversity realized that a community containing a dozen species of the same genus is, in an intuitive sense, less diverse than another also containing a dozen species but with each of them being a member of a different Linnean family. Yet no index of ecological diversity can respond to that distinction as they are all blind to the particular species associated with the measured relative abundances – they are "similarity insensitive" in the terminology of Leinster and Cobbold (2012). Beyond that basic logic, Warwick and Clark (1995) noted that grossly stressed aquatic communities tend to comprise guilds of closely related species, suggesting that the ability to thrive under a particular stressor often occurs within a genus or a family, rather than within a few species distributed across many higher taxa. Any measure of diversity that gives greater weight to distantly related species than to closely related ones will therefore be more sensitive to the stressor, since such a measure will respond to a narrowing in the variety of higher taxa represented in the community. That, however, is to use phylogenetic distance as a proxy for trait differences among species (cf. ICES 2010; Vellend *et al.* 2011) and even greater sensitivity might be achieved by measuring functional trait diversity directly (see Section 5.4).

One approach to quantifying the effects of taxonomy on diversity was to calculate genus diversities (and perhaps family and order ones), standing alongside species diversities and using the same indices (e.g. Magurran 1988). That, however, risked generating a plethora of numerical values which might do more to confuse than enlighten. More sophisticated alternatives, each looking at the phylogenetic relatedness of the species in the community, have therefore been developed[52]. They require the prior existence of a "tree" containing the species of interest and specifically a "rooted tree" – meaning that the species are members of a monophyletic taxocene and that the tree extends far enough to link all of them together. In practice, it is often necessary to limit the approach to taxocenes no larger than a single Linnean class, or even an order, since inconsistencies amongst taxonomists lead to trees that are non-comparable amongst the members of higher taxa (ICES 2010).

---

[52] Ricotta (2005) chose to term all such measures as "weak diversity indices" on the grounds that they violate the axiom that, for a given richness, diversity is maximal when every species in the community is equally abundant. That, however, is not a mark of "weakness" but confirmation that the indices of phylogenetic diversity transcend one of the principal deficiencies of ecological diversity by accepting that, when determining how diverse a community may be, some species are more equal than others.

For most of the biota of Canadian waters, protists and microbiota excepted, taxonomic knowledge is sufficiently mature that each species can be placed in a Linnean tree, branching down from phylum to species[53]. Diversities founded on such trees are sometimes distinguished as "taxonomic diversities" (e.g. Chao *et al*. 2010). Where the necessary data exist, however, it is preferable to use a tree built from the genetic data of molecular taxonomy or, failing that, one constructed on the phenotypic differences of numerical taxonomy. Diversities with such foundations are termed "phylogenetic". In a tree based on classical taxonomy, all of the "links" or "branches" between the "nodes" (or branching points) are considered to have equal lengths. Thus, any three genera in one family are treated as equally dissimilar from one another and as dissimilar as are any two families in one order. Some phylogenetic trees are similarly constructed but others have branch lengths proportionate to observed genetic or phenotypic differences, while in yet others those differences are converted to durations of evolutionary time between branching events and the branch lengths calibrated accordingly.

Vellend *et al*. (2011) have recently reviewed some of the available methods for reducing the diversities of species represented in such trees into scalar index values. They classified the approaches into ones which generate scores for each species in a regional fauna or flora, within the context of that fauna or flora, and then accumulate those scores for such of the species as occur in the community of interest (their "Type I"), *versus* those approaches which consider only species in a particular community, ignoring other regional species ("Type II"). Perhaps the simplest of Vellend *et al*.'s (2011) "Type I" measures is "Taxonomic Distinctiveness", which is the sum (across all species in the community) of the inverse of the number of nodes between a species and the root of the taxocene's tree. For example, a species in a monotypic genus would have one less node than a species, within the same family, that was a member of a genus containing two or more species. Hence the member of the monotypic genus would contribute a higher score to the sum for any community in which it occurs. Because of the summation, Taxonomic Distinctiveness increases with species richness and that is generally true of "Type I" measures, though it would be possible to use mean scores to avoid that dependence (Vellend *et al*. 2011). "Type I" measures are typically used by conservation biologists when prioritizing communities for planning purposes. They allow greater priority to be given to a species that is the sole example of a monotypic genus, for example, than to one with a congener that could replace it if necessary. Once the scores for the species in a particular community are summed, "Type I" measures become metrics of biodiversity but a biodiversity weighted by the (regional, not global) taxonomic uniqueness of the species in the community – the weightings being sensitive to arbitrary definitions of regional boundaries. Such a measure has value in some circumstances. Certain tropical nations, faced with enormous species richness but very small conservation budgets, must prioritize carefully and yet must do so using simple and cheap indices, such as the "Type I" measures. Canadian ecosystems are generally much less diverse, while budgets are rather less restrictive, and hence all species meriting special conservation can receive it. Thus, Vellend *et al.*'s (2011) "Type I" is of limited utility here.

---

[53] Throughout this section, trees based on classical, Linnean taxonomies are described in terms of the primary taxonomic levels (phylum, class, order, family, genus and species) but they could just as well be prepared using the super-, sub- and infra- levels in addition, provided that those are used consistently across all clades in the tree.

Their "Type II" measures, which consider only the taxonomic relationships amongst the members of the community of interest, are typically used by ecologists seeking to understand the composition of communities (Vellend *et al.* 2011) and correspond more closely to metrics of ecological diversity. Some of them assume that all branch lengths in their foundational trees are equal, while others are able to use the information in trees with variable-length branches. Within each sub-type, there are metrics that consider only species' presences but each has an analogue that uses information on the relative abundances of the species. Vellend *et al.* (2011) examined four presence-based "Type II" metrics (Faith's Phylogenetic Diversity, Webb's Mean Phylogenetic Distance, Crozier's Sum of Phylogenetic Distances and Webb's Mean Nearest Neighbour Distance), along with the abundance-weighted alternative to each of those. As an example, Faith's index is the sum of the branch lengths in the phylogenetic tree that link species in the community being evaluated. If a response to relative abundances is desired, the branch lengths are weighted by the average abundance of the species sharing that branch. In practical applications of these "Type II" measures, it is normal to calculate a model-based "null" value that assumes an absence of phylogenetic structure in the community and to express the measured value of the index as a deviation from that standard, which step also serves to remove any dependence on richness (Vellend *et al.* 2011). Vellend *et al.* (2011) went on to examine the behaviour of their eight "Type II" metrics (the four presence-based ones, plus their abundance-weighted analogues), using a simulation approach. From the results, they cautioned against employing any of the eight with trees having even-length branches (thus precluding trees based on classical Linnean taxonomy) and recommended standardization against "null" models but did not offer any preference among the metrics themselves.

One particular pair of metrics that has seen much use in marine applications is Warwick and Clarke's (1995) "Taxonomic Diversity", Δ, and "Taxonomic Distinctness", Δ^*. Vellend *et al.* (2011) regarded them as the abundance-weighted analogue of Mean Phylogenetic Distance, and hence as "Type II" measures capable of using information on branch lengths, but the prominence of Δ and Δ^* in aquatic studies calls for additional attention here.

Warwick and Clarke (1995) saw their Taxonomic Diversity as functionally similar to the Shannon index. It is:

$$\Delta = \frac{\sum\sum_{i<j} w_{ij} p_i p_j}{\sum\sum_{i<j} p_i p_j + \sum_i p_i (p_i - 1)/2} = \frac{\sum\sum_{i<j} w_{ij} p_i p_j}{N(N-1)/2}$$

where w_{ij} is the "distinctness weight" of the path linking the i^{th} and j^{th} species in the tree and N is the total number of individuals in the sample, summed across species (Warwick & Clarke 1995; Clarke & Warwick 1998). If the taxonomic structure is ignored (i.e. all w_{ij} are set to unity), Δ collapses into the Gini-Simpson diversity index divided by $(1 - N^{-1})$ (Clarke & Warwick 1998).

Dividing Δ by what Vellend *et al.* (2011) termed a "null" model yields Taxonomic Distinctness:

$$\Delta^* = \frac{\sum\sum_{i<j} w_{ij} p_i p_j}{\sum\sum_{i<j} p_i p_j}$$

If Δ were weighted by $(1 - N^{-1})$, to produce a metric directly comparable to the Gini-Simpson index, the weights would cancel out of Δ^*, leaving its equation unchanged. Hence, Δ^* can be regarded as the null-corrected, phylogenetically-weighted generalization of the Gini-Simpson index. Meanwhile, the numerators of Δ and Δ^* are the same as Rao's quadratic entropy, a metric sometimes suggested as an index of functional diversity (Weiher 2011).

Both Δ and Δ^* show curvilinear behaviour with increasing sample size but, as would be expected from the relationship to the Gini-Simpson index, reach their asymptotic values with rather fewer samples than are required to achieve the same with Shannon's H' and many fewer than are needed for species richness[54]. As is also to be expected, Δ varies with S but Δ^* is independent of it (Warwick & Clarke 1995)[55].

Warwick and Clarke (1995) tested Δ and Δ^* with monitoring data gathered around the Norwegian *Ekofisk* oil field, where the drilling had used toxic diesel-based muds and the oiled cuttings had been discharged to the seabed. Measures of macrobenthic ecological diversity, including species richness, ecological diversity *sensu stricto* (Shannon's H') and evenness (probably J', though described as "Pielou's J"), were depressed very near the cuttings pile but reached background levels a few hundred metres away. In contrast, Δ and Δ*, calculated from a tree representing classic Linnean taxonomy, continued to increase with distance from the pile out to the furthest sampling stations, at several kilometres range (Warwick and Clarke 1995). Even multivariate ordination had only detected effects to 3 km around *Ekofisk* (Gray *et al.* 1990). Further experience over the following decade has supported that early positive impression of Δ and Δ* (ICES 2010), though Hall and Greenstreet (1998) found that long-term trends in Δ and Δ* of North Sea groundfish closely followed ^1D and ^2D, calculated from the same data.

---

[54] Magurran (2004) stated that Δ^* shows a "lack of dependence on sampling effort". The lack is incomplete, though Δ^* is less dependent on N than is H'.

[55] Leonard *et al.* (2006) have represented that independence of S as though it were independence of sample size. It is not, despite the dependence of S on N.

For presence / absence data, Clarke and Warwick (1998) further simplified the index into "average taxonomic distinctness":

$$\Delta^+ = \frac{\sum\sum_{i<j} w_{ij}}{S(S-1)/2}$$

where S is the sample (or pooled sample) richness. That metric is simply the average of all possible pairs of w_{ij}, with $i \neq j$, though calculated from one half of those pairs. Δ^+ will be encountered later as the foundation of an index of functional diversity. It may be noted that this metric was designed to be independent of S, while it ignores the relative abundances of the species. Thus, while it can be regarded as an index of some facet of biodiversity (linked to the amount of taxonomic variability within a community), Δ^+ is divorced from the foundations of ecological diversity, in contrast to Δ.

Stressed environments might be expected to show groups of closely-related species, with small numbers of distantly-related ones, that could have values of Δ^+ identical to those in more taxonomically-balanced communities, since the index has the nature of an average. Clarke and Warwick (2001) therefore suggested a further measure, "variation in taxonomic distinctiveness", that quantifies the variability within the matrix of w_{ij}:

$$\Lambda^+ = \frac{\sum\sum_{i \neq j} w_{ij}}{S(S-1)} - \left(\Delta^+\right)^2$$

Clarke and Warwick (2001) correctly emphasized the independence of that measure from S and hence from sampling effort but risked confusing users of Λ^+ by declaring that that independence justifies use of the metric in comparing species lists prepared by different workers – a confusion that can be seen in the work of Leonard *et al.* (2006), Bevilacqua *et al.* (2011) and others. Independence from quantitative differences in sampling effort is not the same as independence from qualitative differences in sampling protocols and valid application of Λ^+ is as dependent on standardization of survey methodology as is that of any other metric.

Some rather exceptional claims have been made for these four metrics, Δ, Δ^*, Δ^+ and Λ^+, as measures of anthropogenic impacts. Leonard *et al.* (2006), for example, claimed that Δ^+ can be used with uncontrolled sampling (e.g. following different protocols), even remaining unaffected by whether the data collection does or does not utilize taxonomic expertise; that it is independent of sampling effort; and that it can provide valid comparisons across habitat types, being insensitive to natural spatial variation, while remaining sensitive to anthropogenic change – including change resulting from such contrasting drivers as heavy-metal contamination and seabed disturbance by beam trawls[56]. Clarke and Warwick (1998), working with data on free-living marine nematodes, proposed that 95% confidence limits for Δ^+ could be simulated by modelling data using an assumption that the species compositions seen in individual samples are random selections from the regional species pool – which incorporates an improbable assumption that all of the members of the pool have equal frequencies of occurrence. Clarke and Warwick (2001) later extended that analysis to include Λ^+. However, when Somerfield *et al.* (2008) applied Δ^+ to data from groundfish surveys, they found a major bias resulting from among-species differences in frequencies of occurrence, necessitating different assumptions underlying the simulation of confidence limits. That may be a suitable caution against relying on either version. Meanwhile, Clarke and Warwick (2001) observed differences in the values of Λ^+ between supposedly "pristine" and degraded communities, judging their significance from the confidence limits. It is clear, however, that Δ^+ varies naturally as well as being depressed by some kinds and magnitudes of anthropogenic stress (Clarke & Warwick 1998). The dual causes of variations preclude inferences about causation, unless the metrics are applied within an appropriate replicated and randomized experimental design.

Some subsequent authors have found value in Clarke and Warwick's (1998, 2001) metrics (e.g. Xu *et al.* 2011, 2012). In contrast, Salas *et al.* (2006) applied those measures to various data sets from Iberian estuarine, lagoonal and intertidal environments, finding that Δ^+ and Λ^+ were relatively insensitive to anthropogenic degradation, when compared to other indices such as *H'*. The taxonomic distinctiveness measures did not show the claimed monotonic relationship to the degree of degradation. Abellán *et al.* (2006) found much the same in a riverine environment. Bevilacqua *et al.* (2011) tested the metrics using data collected around an offshore gas production platform in the Mediterranean Sea and another dataset from the vicinity of a sewage outfall. In both cases, metrics sensitive to *S* showed anthropogenic effects but Δ^+ and Λ^+ did not, though they did vary among habitats. Most recently, Bevilacqua *et al.* (2012a) applied the metrics to Mediterranean nematodes, leading them to reject the suggestion that Δ^+ is invariant across natural systems. They concluded that it is a useful tool for exploring biodiversity patterns within larger assessments but that it is not a panacea for detecting anthropogenic disturbance. That seems a fair summary.

---

[56] Leonard *et al.* (2006) were able to illustrate those claims with particular datasets, at least while they were free to interpret all observed changes in Λ^+ as anthropogenic in origin. It seems unlikely, however, that they would stand up to more extensive testing. In particular, if the index values are to respond to input data at all, different sampling intensities or protocols in different locations must lead to differences in Λ^+ that are purely artifacts of "uncontrolled" sampling. Meanwhile, different anthropogenic stressors have different effects on biotic communities, as do different axes of natural variation. It is grossly improbable that all of the anthropogenic change would result in a depression in the average taxonomic distance between observed species that could be unambiguously distinguished from natural differences.

In a contribution too recent for inclusion in Vellend *et al.*'s (2011) review of phylogenetic diversity indices, Chao *et al.* (2010) linked the idea to Hill Numbers. They showed a failure of existing indices to respond as one would expect diversity to do, analogous to the failures that Jost (2006) had shown for many metrics of ecological diversity. In their place, Chao *et al.* (2010) offered:

$${}^{q}\overline{D}(T) = \left\{ \sum_{i \in B_T} \frac{L_i}{T} a_i^q \right\}^{1/(1-q)}$$

where L_i is the length of branch *i* in the set $\boldsymbol{B_T}$, containing all branches over the time period *T* years preceding the present, a_i is the summed relative abundance of the species descended from branch *i*, *T* is the time since the root of the tree or other node that may be of interest, and *q* is the order of the measure. If all species were maximally distinct, meaning that they had been distinct from one another since the root of the tree, all L_i would equal *T* and the above equation would collapse into that defining ${}^{q}D$. With real taxonomies, ${}^{q}\overline{D}(T)$ is equal to the number of equally abundant and maximally distinct species in a hypothetical community that has phylogenetic diversity equal to that of the real community under consideration, in the same way that ${}^{q}D$ is equal to the "effective number of species" of "true" ecological diversity. If the foundational tree was drawn based on genetic or phenotypic differences, rather than evolutionary time, the same equation can be used, though with *T* replaced by whatever measure of difference was used in preparing the tree (Chao *et al.* 2010).

Chao *et al.* (2010) went on to show how their ${}^{q}\overline{D}(T)$ is related to transformations of other indices of phylogenetic diversity. ${}^{0}\overline{D}(T)$ is equal to Faith's Phylogenetic Diversity measure (in its original form, without weighting by species' relative abundances) divided by *T*. ${}^{1}\overline{D}(T)$ equals the exponent of Allen's Phylogenetic Entropy:

$$H_p = -\sum_i L_i a_i \log a_i$$

which is a generalization of the Shannon index. Finally, ${}^{2}\overline{D}(T)$ is equal to $T / (T - Q)$, where Q is Rao's quadratic entropy:

$$Q = \sum_{ij} d_{ij} p_i p_j$$

in which d_{ij} is the phylogenetic distance between species *i* and *j*, and p_i is the relative abundance of species *i* – a metric related to Warwick and Clarke's (1995)

Δ, both being regarded by Vellend *et al.* (2011) as forms of Mean Phylogenetic Distance, and hence having much in common with the Simpson index.

Chao *et al.*'s (2010) ${}^{q}\overline{D}(T)$ has all of the advantages of ${}^{q}D$ in its proportionality to expectations of the way that diversity should respond to shifts in communities, as well as the flexibility of adjusting sensitivity to rarities by altering q. The close connections between ${}^{q}D$ and ${}^{q}\overline{D}(T)$ should also aid comprehension of the sometimes-complex concept of phylogenetic diversity. The behaviour of Chao *et al.*'s (2010) metric has yet to be extensively investigated, however, and thus cannot be wholeheartedly recommended at this time, though it does appear promising.

Yet another alternative has recently been offered by Leinster and Cobbold (2012), who also built on ${}^{q}D$ and derived an expression for an "effective number of species" in which those "species" are maximally distinct. Their approach, however, is not tied to taxonomic or phylogenetic diversity but can accommodate measures of any form of similarity (or difference) among species – meaning that it is, as they termed it, a "similarity sensitive diversity measure" (Leinster & Cobbold 2012). Leinster and Cobbold (2012) suggested, for all species with non-zero abundances and for $q \neq 1$ or ∞:

$${}^{q}D^{Z}(p)=\left\{\sum_{i=1}^{S}p_i(Zp)_i^{q-1}\right\}^{1/(1-q)}$$

which is the "similarity insensitive" Hill Number weighted by:

$$(Zp)_i=\sum_{j=1}^{S}Z_{ij}p_i$$

in which the Z_{ij} are the elements of an $S \times S$ matrix containing measures of similarities (each in the range zero to one) among the species. Equations for the limit values of ${}^{q}D^{Z}(p)$ when $q = 1$ or ∞ were provided by Leinster and Cobbold (2012). Necessarily $(Zp)_i \geq p_i$ and hence ${}^{q}D^{Z}(p)$ is almost always smaller than ${}^{q}D$, the diversity of a real community being one of fewer "effective maximally distinct species" than of "effective species" some of which are similar to one another.

Leinster and Cobbold (2012) showed that, for $0 \leq q \leq \infty$, Chao *et al.*'s (2010) ${}^{q}\overline{D}(T)$ is equal to ${}^{q}D^{Z}(p)$ when applied to an "ultrametric" tree (one scaled such that the lengths from root to tip are equal for all tips). For non-ultrametric trees, Leinster and Cobbold (2012) had to invoke "relatedness matrices", a class that incorporates but is broader than similarity matrices. With that modification, their ${}^{q}D^{Z}(p)$ is equal to Chao *et al.*'s (2010) ${}^{q}\overline{D}(T)$ for all trees, though the former is more adaptable as it can be applied with any similarity measure.

${}^{2}D^{Z}(p)$, the similarity-sensitive analog to the Inverse Simpson metric, reduces to the inverse of the expected similarity between any pair of randomly chosen individuals in the community. Indeed, for all integer $q \geq 2$:

$$^{q}D^{Z}(p) = \left\{\mu_q\right\}^{1/(1-q)}$$

where μ_q is the expected similarity among the members of any group of q individuals randomly chosen (with replacement) from the community (Leinster & Cobbold 2012). That can be a valuable property in some situations, such as when working with microbial diversity, in which similarities among individuals can often be quantified even though the divisions into species are poorly known.

Leinster and Cobbold (2012) stressed that their approach can be used with any measures of similarity that are relevant to the biological issue at hand, mentioning genetic and functional similarities. It could clearly be applied to taxonomic similarity. That very adaptability, however, while making $^{q}D^{Z}(p)$ a powerful research tool, limits its use in routine monitoring. If approaches to quantifying similarity can be standardized, routine use might become sensible but, until then, the results obtained will be heavily dependent on analysts' preferences for the types of similarity considered and the indices of similarity adopted. Furthermore, Leinster and Cobbold (2012) developed their mathematics for fully-censused communities. Appropriate ways of estimating $^{q}D^{Z}(p)$ from samples have yet to be addressed. Hence, theirs is a promising approach for the future but not one for immediate application outside of research contexts.

The question remains of whether or not it would be beneficial, in routine work in support of management, to supplement measures of ecological diversity with some index of phylogenetic or taxonomic diversity. The latter certainly utilize more of the information in a set of species-abundance data, though the difference should not be exaggerated: Jost (2008) and Chao *et al.* (2010) have termed the metrics of ecological diversity "species-neutral" and contrasted that with the indices of phylogenetic diversity, which they termed "non-neutral measures", but those seem unfortunate terms. The quantification of ecological diversity for any one community would be unaffected if the species were randomly ordered and labelled A, B, C ..., while the assignment of those labels to particular species need bear no relation to the A', B', C' ... labels given to the species in some other community, even when the diversities of the two communities are being directly compared. It would be an over-simplification to say that taxonomic diversities could be calculated from data similarly labelled as A_i, A_{ii}, A_{iii}, B_i, C_i, C_{ii} ..., where the letters represent genera and the subscripts species within those genera. That would be an over-simplification because a more complex structure incorporating labels for phyla, classes, orders and families, as well as genera and species, would be needed. However, the principle that the indices could be calculated from taxa that have been randomly assorted within their next-higher taxon and given arbitrary labels remains true, as does the independence among communities of the randomization and labelling. Even Leinster and Cobbold's (2012) approach considers only the similarities amongst the species present within each community, disregarding any commonality (or lack thereof) amongst the species lists of compared communities. Thus, phylogenetic and taxonomic diversity indices utilize more of the information in a dataset than do the metrics of ecological diversity but still far less than could be captured by multivariate analyses that preserve the identification of relative abundances to particular, named species.

Elsewhere in this report (see Sections 2.5 & 4), we recommend that values for ecological diversity *sensu lato* be routinely reported not because they are meaningful (which they generally are not) but because there is a broad expectation amongst non-specialists that ecological diversity is the core of the CBD's biodiversity. Phylogenetic diversity cannot substitute for ecological diversity in that role because it does not carry the same expectation and cannot readily be explained to non-specialists. Metrics such as Δ^*, Δ^+, Λ^+, ${}^q\overline{D}(T)$ or ${}^qD^Z(p)$ could be used alongside those of ecological diversity but that is unlikely to be a primary choice – multivariate statistical approaches offer more power and flexibility. Thus, where quantification of Canadian aquatic biodiversity is concerned, phylogenetic (and taxonomic) diversity lacks much of a role.

If measures of this form of diversity were to be employed, their use should be preceded by a research effort, essentially repeating Vellend *et al.*'s (2011) trials but with particular reference to the types of datasets available for Canadian aquatic biota and incorporating a wider variety of "Type II" metrics, especially ones that use information on relative abundances and including those of Chao *et al.* (2010) and Leinster and Cobbold (2012), which well merit such testing but require it before they are used routinely. If Vellend *et al.* (2011) then prove correct that "Type II" indices can be seriously misleading when applied to taxonomic trees or to phylogenetic trees with equal branch lengths, there will be a major difficulty in quantifying these diversities. Trees based on well-developed molecular phylogenies or numerical taxonomies are available for some clades but probably not for many entire taxocenes found in Canadian waters, while even the available trees are based on particular genetic markers – different markers typically yielding different trees. Where adequate trees do not exist, it is doubtful whether the cost of preparing them would be justified by their use in measuring phylogenetic diversities. Hence, even experimental application of the metrics, in a research setting, is likely to be confined to those taxocenes for which a molecular phylogeny has been prepared for quite other purposes.

5.4 FUNCTIONAL TRAIT DIVERSITY

Functional diversity is the diversity of traits displayed by the species in a community. It not only preserves the information in the species' identities, as captured in their names (thus going beyond the constraints of phylogenetic diversity), but allows consideration of all that is known of the autoecology associated with the name.

In some ways, functional diversity is what ecological diversity should have developed into. If two congeners speciated during a period when their populations were allopatric but have since occupied one another's ranges, while maintaining a taxonomic distinction, then the two will be functionally interchangeable (essentially indistinguishable) in ecological terms. From a human perspective, finding attractiveness in variety, the existence of both of the two distinct species might be an important element of biodiversity but, ecologically, it is no more relevant than an equivalent amount of intra-specific genetic diversity would be. However, should the two species have developed different traits, in the sense that Darwin's finches have evolved different bill shapes and not merely contrasting colours in the feathers used to attract mates, then the existence of two different species becomes a matter of ecological relevance. In essence, ecosystem processes are not affected by two congeners having been given different names but by the functional differences between species (Ricotta 2005). Hence, the use of diversity in ecological research should not have been focused on the relative abundances of species *per se*, nor even on species placed within taxonomic hierarchies, but rather on species that carry distinct traits and on the degree of difference among the traits of different species. Where a relationship can be shown between ecological diversity and some community property, such as productivity, it comes through the functional diversity of the species whose abundances have been captured by the index of ecological diversity (see Section 2.3). Even community stability, which seems related to redundancy, requires species that can replace one another and yet are sufficiently functionally different that they do not respond to the same stressors. To the extent that phylogenetic diversity provides a more sensitive indicator of anthropogenic change in communities than ecological diversity does, it is because phylogenetic similarity tends to be correlated with similarity in functional traits.

Furthermore, modern ideas of functional diversity do not have their roots in attempts to summarize species-abundance data into a single index, as ecological diversity and even phylogenetic diversity do, but rather in early studies of the ways that ecosystems work. Hence, instead of boiling down expensive data into a single scalar value, the approaches developed for examination of functional diversity emphasize multivariate analyses, distances in hyper-dimensional "trait space" and other complexities (Weiher 2011). That is what is needed in modern ecological research, including in the analyses of Species Abundance Distributions, but it renders functional diversity into a quite different topic than ecological diversity as the latter is generally understood.

Weiher (2011) has defined "functional diversity", which he considers almost synonymous with "trait diversity", as the degree to which co-existing species vary in their functional traits, the latter being those observable (or operationally defined) phenotypic characteristics that influence species performance and/or ecosystem processes. Many approaches have been used for quantifying and examining functional diversity but their application remains challenging and better suited to specific research studies than for monitoring achievement of policy goals. Even selecting which traits to consider is complex and not amendable to generating uniquely ideal outcomes. After that selection, the chosen traits must be scaled so that the necessarily different units of original measurement (e.g. lengths, areas and speeds) do not distort analyses. Weiher (2011) warned that there is no current consensus on the best way to do that. Overall, he concluded that there are "few, if any, bona fide standard practices" in quantifying functional diversity. Ricotta (2005) went further, suggesting that "a set of basic statistical requirements that a meaningful index of functional diversity should satisfy has never been suggested".

It has sometimes been proposed that functional diversity should be quantified using metrics of phylogenetic diversity, substituting similarity of traits for taxonomic relatedness (e.g. Somerfield *et al.* 2008; Chao *et al.* 2010; Leinster & Cobbold 2012). That approach cannot be recommended, however. We would instead point to "traits analysis", a multivariate approach not unknown to ecologists (e.g. Kenchington *et al.* 2007) that allows joint ordination of multiple, linked matrices. The common forms of ordination (e.g. PCA or nMDS) work on a single two-dimensional matrix, such as one containing the catches of multiple taxa at each of many stations. They can show the relationships among those taxa as they are expressed across the stations – which may be spread across time as well as space. Traits analysis, in contrast, can combine a species x stations matrix with a species x traits matrix, revealing how the traits are expressed across the stations, while the species fall out of the analysis. If desired, a third matrix could be added, perhaps stations x environmental-characteristics, to examine how the traits are related to such features as depth, temperature or proximity to point sources of anthropogenic pressures. Should there be reason to think that phylogenetic relationships contain more information on the ecological function of the species than do observable traits, then those relationships can be coded into the species x traits matrix. Thus, traits analysis offers a far more flexible and powerful alternative to the over-simplifications of any scalar metric and one that has been proven in research settings. What is now required is development work exploring its application to routine measurement of biodiversity.

5.5 BIOTIC INDICES

While they are not measures of biodiversity, in the sense of the CBD definition, another class of useful indices related to functional traits merits mention here. That class has been given various labels, including "Index of Biological Integrity" ("IBI"), "Marine Biotic Index" ("MBI") and "Benthic Quality Index", though such measures were known as "Biotic Indices" as early as a review by Washington (1984) and that usage will be followed here. Typical biodiversity metrics capture information on spatial or temporal variations in the biota, from which inferences may be drawn on the quality of the environment without requiring knowledge of the biology of the species present – indeed, often ignoring which species those may be. In contrast, the biotic indices rely on the known responses of particular species to particular anthropogenic stressors. They are founded on recognition of some gradient of environmental degradation. Species are then scored according to their tolerance of the pollutants or disturbance in question, the scores for the various species are combined into a single metric, which is validated against independent data. Thereafter, degrees of environmental degradation elsewhere, or in the same location at a later time, can be assessed based on observations of the species occurring in an area, used as surrogates for direct measures of physical or chemical conditions – often with the advantage that the biota integrate the effects of anthropogenic pressures over their lifetimes. Soft-bottom benthic communities are commonly chosen for such work because of the ease of sampling, the lack of long-distance movement of individuals within those communities (ensuring that the measured signal is specific to the place where the samples were taken), the tendency for much that is in the water column, including suspended chemical pollutants, to be deposited in areas of fine sediments, and the vulnerability of sedentary benthic species to any local hypoxia events (e.g. Weisberg *et al.* 1997; Borja *et al.* 2000, 2003, 2008, 2009; Rosenberg *et al.* 2004; Muniz *et al.* 2005; Muxika *et al.* 2005; Borja & Dauer 2008).

Biotic indices have merit in monitoring and mapping what is here termed environmental status but they do not directly address variability in biological systems. Washington (1984) stressed that they do not measure community structure and hence are not metrics of what is here termed ecological diversity. Thus, all such indices fall outside the scope of the present report and we only address them briefly to explain what might otherwise appear to be a *lacuna* in this overview of biodiversity metrics.

Many biotic indices have been developed, for application in different areas or to different types of anthropogenic degradation – Diaz *et al.* (2004) were able to list 64 variants applicable to various benthic communities, Marques *et al.* (2009) reviewed 31 suited to estuarine and coastal ecosystems[57] and further development continues (see Borja & Dauer 2008 and following papers; Borja *et al.* 2009 and following papers). Probably the currently most advanced example, and the only one recommended by Marques *et al.* (2009) after practical applications to some Iberian embayments, is "AZTI's Marine Biotic Index" ("AMBI"), which was originally developed for European inshore waters. It uses a list of some 3,000 soft-bottom benthic species which have been classified into five groups:

GI: Species very sensitive to organic enrichment, scored zero,
GII: Species indifferent to enrichment, scored 1.5,
GIII: Species enhanced by enrichment, scored 3,
GIV: 2nd order opportunistic species adapted to unbalanced conditions, scored 4.5, and
GV: 1st order opportunistic species adapted to pronounced unbalanced conditions, scored 6.

---

[57] Marques et al. (2009) also reviewed a dozen biodiversity metrics. Unfortunately, their understanding of that topic was seriously deficient and their remarks on it are best disregarded.

The proportion of total sample biomass that falls into each group is multiplied by that group's score and the five results summed into a value that must lie between zero and six. Results of <1.2 are deemed "normal", >3.2 but <5 is taken to indicate "moderately polluted", while 6 is "very highly polluted" (Marques *et al.* 2009). AMBI has a multivariate, or more exactly a multimetric, extension, "M-AMBI", which uses factorial and discriminant analyses to combine the AMBI score with species richness and Shannon's *H'*, thus avoiding a need to report multiple indices separately (Marques *et al.* 2009). The general approach was developed for anthropogenically-stressed lacustrine, estuarine and even coastal environments. It has proven to be valuable in some ecosystems and it has been widely applied in Europe, in North American freshwaters and elsewhere.

While the use of information on the biological responses of species makes biotic indices potentially more powerful than any metric of biodiversity could be, it also means that any one biotic index can only be used within a particular habitat type (e.g. soft-bottom, subtidal estuarine areas: cf. Diaz *et al.* 2004), sometimes only within a particular sampling season when certain species are available to the survey gears (Borja & Dauer 2008), and only within the geographic ranges of its indicator species (Washington 1984) – though AMBI is adaptable and has been successfully used outside Europe by adding local species to the lists for each of the five groups (e.g. Muniz *et al.* 2005; Borja *et al.* 2008; Borja & Tunberg 2011; Teixeira *et al.* 2012). Moreover, any one biotic index can only address the particular suite of anthropogenic pressures for which it was developed (cf. Washington 1984), the preminence of which in a study region must be determined *a priori*. Thus, AMBI failed in the Baltic Sea, where strong salinity gradients shape benthic communities, overwhelming anthropogenic factors (Zettler *et al.* 2007). Indeed, all estuarine ecosystems are stressed by the moderate and variable salinities, while many are also notable for their high levels of organic inputs – making it difficult to distinguish anthropogenic stress or eutrophication from natural conditions (e.g. Dauvin & Ruellet 2009). As Marques *et al.* (2009) summarized their conclusions from practical applications of what they termed "ecological indicators":

> despite their potential utility, most ecological indicators are more often than not relatively specific for a given kind of stress, or applicable to a particular type of community and/or scale of observation[.]

We would disagree only to the extent of replacing "or" with "and". The primary forms of anthropogenic degradation endured by estuarine waters of western Europe or the eastern seaboard of the United States may be those arising from urbanization and industrialization, which perhaps produce a sufficiently consistent suite of pressures that their effects can be measured using a single index, such as AMBI. Along undeveloped coastlines, such as most of those in Canada, or in offshore waters, quite other anthropogenic pressures will predominate and the resulting changes in biotic communities are expected to be qualitatively, not merely quantitatively, different from what is seen in urbanized estuaries, making it inadvisable to use an biotic index imported from Europe.

More generally, the measurement of some facet of biodiversity, followed by the drawing of inferences, is an exploratory process that need not rely on *a priori* assumptions about anthropogenic pressures nor the sensitivity to them of various species. Applying a biotic index, in contrast, requires prior knowledge of both and hence is only suited to very well-studied systems. If applied incautiously elsewhere, too close a focus on any one such index could result in major environmental degradation, of an unanticipated kind, passing unrecognized. Where they are used, biotic indices should be regarded as measures of ecosystem responses to particular stressors, rather than metrics of overall environmental status.

5.6 HABITAT AND ECOSYSTEM DIVERSITY

The diversity of habitat or ecosystem types within spatial units is an important facet of biodiversity but means for measuring it have not progressed very far. The most common alternative appears to be preparing an ecosystem classification, mapping occurrences of the various recognized types and then applying the metrics of ecological diversity, presumably using units of percentage cover as the measure of "abundance". That approach presumes, however, that it is possible to uniquely classify each small area into a discrete habitat or ecosystem type, which is rarely possible. Perhaps Leinster and Cobbold's (2012) ${}^{q}D^{Z}(p)$ could be applied, using information on the habitat similarities amongst randomly selected points within the area of interest, though no such development has yet been attempted. Otherwise, β diversity may be useful as an indirect measure of the diversity of habitats based on the diversity of communities.

Mapping occurrences and extents of particular habitat types can be more straightforward, though not necessarily easy, and may provide a useful indicator of environmental status if the map is of exceptionally valuable or vulnerable types. That approach does not, however, address variability and hence cannot measure a facet of biodiversity.

5.7 SIZE COMPOSITIONS AND SIZE SPECTRA

Size compositions within a population represent variation amongst individuals and hence are a facet of biodiversity. Moreover, changes in those compositions can be an important indication of trends in environmental status, particularly for species with indeterminate growth and hence sizes linked to ages – as is true of most aquatic animals and macrophyte plants. Since size differences (whether within or among species) are also related to variations in trophic level in many marine organisms, size compositions carry important information about ecosystem structures and functions. Size spectra can be linked to models of ecosystem fluxes and hence to the magnitude of ecosystem services (Dornelas *et al.* 2011). Finally, in offshore ecosystems size compositions are among the most sensitive of all aspects of biodiversity to anthropogenic pressure: the direct effect of fishing effort being an increase in mortality rates, which decreases life expectancy and hence the time available for the average individual to grow – an effect that is commonly exacerbated by targeting practices, which tend to favour the harvesting of larger fish, and by the commonly lower ability of large animals to sustain any given level of fishing mortality, itself a consequence of size-related differences in natural mortality rates (cf. Shin *et al.* 2005; Greenstreet 2008). Monitoring of individual sizes is a well-established practice in fishery resource surveys and there is an attraction in using the resulting size information in biodiversity monitoring (cf. Rice 2000, 2003). Unfortunately, in practice those data cannot support highly-meaningful measures[58].

More than a dozen indices based on individual sizes, termed "Size-Based Indicators" or "SBIs" (Shin *et al.* 2005) have been suggested, including both simple ones, such as the community-wide mean length of the members of some taxocene, and others more complex. The latter include the slopes, intercepts and curvatures of size spectra, drawn in either length or weight units (ICES 2005; Shin *et al.* 2005; Charnov *et al.* 2012). Most of those measures have been shown to respond to fishing pressure as they are expected to do (e.g. mean size falling as fishing effort increases), though the size spectra tend to be rather insensitive and so-called "diversity spectra" even more so. A much greater deficiency is that the full development of a response in size composition lags many years behind changes in fishing effort – at least one exploited lifetime of the affected species and sometimes much more if the anthropogenic effect involves ecosystem responses to the depletion of the targeted fish (Shin *et al.* 2005)[59]. In consequence, the signal communicated by an SBI tends to reflect management choices made long before, rather than those currently in effect. Poor outcomes are to be expected whenever a control process aims for past positions of a "moving target" and hence SBIs could lead to very serious consequences if incautiously used as a basis for advice to decision-makers.

All SBIs face further severe challenges. Firstly, in most marine species recruitment success is highly variable, resulting in great variability in year-class strengths. In consequence, a downward shift in the sizes of individuals in a population can result either from depletion of larger individuals, potentially driven by anthropogenic factors, or from a strong recruitment event, which would normally have non -anthropogenic causes. An upward shift can be a result of reduced mortality rates on exploited size-classes, a decline in recruitment (including a return to normal levels after a strong recruitment event) or simply the progression to higher age, and so sizes, of a strong year-class that recruited some time before. Greenstreet *et al.* (2011, 2012b) observed just such an increase in their "LFI", driven by a strong year-class of saithe (*Pollachius virens*) growing past a length taken as defining a "large fish". Observed size compositions are also affected, over similar time periods to their responses to fishing effort, by shifts in water temperature and food availability (the latter sometimes but not always having anthropogenic causes) through their effects on species' growth rates and hence on individual sizes, as well as by responses to potential genetic change driven by selective fishing (Shin *et al.* 2005). Moreover, most SBIs yet suggested carry a further weakness in that they compound within-population changes in size compositions with shifts in relative abundances among species that differ in adult sizes.

---

[58] Size information is here discussed only in terms of temporal change. For the sedentary benthos and any other species which do not undertake ontogenetic migrations after their larval phases, spatial variations in individual sizes might be meaningful. For example, closures intended to protect sponge grounds might be designed around areas where the species of interest are able to grow to the largest sizes. We have not found any examples of size data being used for such purposes with reference to marine ecosystems.

[59] These long lags seem to have surprised some users of SBIs (cf. Greenstreet *et al.* 2011; Shephard *et al.* 2011). They should have been expected: If, hypothetically, two periods of stasis in fishing effort were separated by a major step change (either an increase or a decrease), the age composition of the resource, and hence its size composition, would not return to stability until every surviving year-class had known only the new level of effort throughout its exploited lifetime. That process could take decades, depending on the species in question and the post-change rate of fishing mortality.

Given full data on the abundances of each size class in a population, it is sometimes possible to disentangle the various drivers and draw conclusions about the state of that population. In many situations, however, no SBI could be interpreted to provide any unique explanation for its observed temporal changes, while the further that the data are compressed into simpler indices, with the loss of information inevitable in such compression, the more doubtful interpretations will become. Indeed, Shin *et al.* (2005) recognized that "diagnosis of population state is not straightforward" when working with SBIs and that the "main problem is to disentangle the different sources of variation". They were reduced to calling for changes in mean lengths to be interpreted alongside information on condition factor and size-at-age, as well as for SBIs to be supplemented with information on the proportions of large and small species in the community and on the abundances of recruits (Shin *et al.* 2005) – which calls into question the value of summarizing size data into a simple index in the first place.

Nicholson and Jennings (2004) examined the statistical power to detect changes in five SBIs when applied to data from the North Sea International Bottom Trawl Survey series – reckoned one of the best datasets on any European marine ecosystem. They found that reliable detection of realistic rates of change in each of mean fish length, mean individual weight, weighted means of species maximum lengths or weights, or the slopes and intercepts of size spectra would require data from decades of annual surveys – a time scale of limited utility to management (Nicholson & Jennings 2004). To the extent that that conclusion is true of the application of SBIs to other data sets, it suggests that they have little practical value as measures of biodiversity.

In recent years, one particular SBI, the "Large Fish Indicator" or "LFI", has drawn considerable attention, particularly in the United Kingdom though also in such European organizations and initiatives as OSPAR and the EU's Marine Strategy Framework Directive. The LFI has been advanced not simply as a metric of size diversity but also as an indicator of food-web structure (ICES 2011a, 2012b; Greenstreet *et al.* 2011, 2012b; Shephard *et al.* 2011; Fung *et al.* 2012). The LFI is the proportion of the biomass[60] of some selected set of species (as seen in the data from a particular standard survey series) that is represented by individuals larger than a chosen critical size (for groundfish: currently 40 cm in the North Sea, 50 cm in the Celtic Sea and 30 cm in the Baltic, where the index considers data on only four species) – though the biomasses are normally calculated from length data and the length / weight relationships for the species in question, rather than from weights taken at sea, and thus ignore variations in condition factor. The length threshold is usually selected so as to maximize the signal from the index, given the available data. The quantitative value of the LFI is then commonly compared with one calculated for a period during which the regional fishery resources are deemed to have been "sustainably exploited" (itself a dubious notion), apparently based on the results of stock assessments – the availability of which would seem to negate much of the point of a simple, size-based index. As expected of any SBI, the LFI shows lags in the full development of its responses to changes in fishing pressure, lags which have been found to be 12 to 16 years for North Sea and Celtic Sea groundfish. That delay in the measured response has led to suggestions that simulation methods be used to project forwards in time, such that current advice to managers can reflect the effects of recent management (ICES 2011a), though that would remove the value of using empirical observations by substituting model results. As if the basic LFI did not compress too much real complexity into a single, scalar value, there have also been attempts to apply it across multiple surveys of the same region (ICES 2011a).

Unfortunately, that entire effort only yields a measure of the relative abundances (or more exactly biomasses) of individual fish larger and smaller than some size which, while not strictly arbitrary, has no foundation in ecology or conservation biology. That measure is responsive to both within-species changes in size compositions and shifts in the relative abundances of species. Whatever relationship the value of LFI has to anthropogenic pressures is to pressures so far into the past as to have minimal relevance to current management decisions. Those weaknesses are shared by such simple measures as the mean length of fish in a community but the latter does not carry a misleading patina of advanced "scientific" complexity.

We would suggest that, in general, data on individual sizes should not be summarized by any overall metric, like the LFI, but should be examined and interpreted on a population-by-population basis, within comprehensive assessments of those populations – that being the only means by which changes in size compositions can be understood and interpreted in terms of their causative factors (cf. Shin *et al.* 2005). The bulk of the available data is likely to be drawn from resource populations, for which most change in sizes results from shortened life-expectancy caused by adding fishing mortality to the rate of natural mortality. In a Canadian context, that effect is better addressed through stock assessments than as an aspect of biodiversity monitoring, divorced from the detailed information available on the fisheries and their management. Any conclusions from those assessments should be reported as trends in the populations in question, rather than those in the sizes of individuals specifically. Where changes in size compositions merit reporting as part of population-level advice, we recommend using whichever simple summary statistic best illustrates the trends revealed by assessment. That might be mean length, proportion of the population larger than the size at 50% maturity, the length at some selected age, the estimated abundance of individuals larger than 75% of the asymptotic length or any number of other measures. The "Proportional Stock Density" and "Relative Stock Density" measures long used for freshwater recreational fishery resources (e.g. Guy *et al.* 2006) are examples of such measures.

---

[60] Biomass is preferred over abundance because it reduces (though it cannot eliminate) the effects on size compositions of recruitment variations (Greenstreet *et al.* 2011; ICES 2012b). It will also have the effect of weighting the value of LFI towards those species with large adult sizes.

Should a need be seen for community-wide size metrics, for use in illustrating scientific advice, we strongly recommend those which separate within-species and among-species trends. The latter requires averaging of comparable single-population indices, while the comparability demands adjustment of measured sizes to allow for differences in growth patterns – a criterion that was not addressed by any of the indices considered by Shin *et al.* (2005). For fish species, for most of which the parameters of the von Bertalanffy growth equation have already been estimated, the simplest measure would be the within-species proportion of individuals taken in the chosen standard survey that are larger than some set fraction of the asymptotic length, L_∞. A species with isometric growth that follows the von Bertalaffy model exactly (with that model's t_0 parameter set at zero) necessarily shows a point of inflexion of its curve of growth in weight at $^2/_3\,L_\infty$. A common simplifying assumption of teleost life-history theory is that that point is also approximately the size at first maturity and the age at which a year-class would achieve its maximum biomass in the absence of fishing (e.g. Roff 1984; Jensen 1996). Hence, that is a convenient marker separating "large" and "small" individuals of any one species, and yet enjoys some biological foundation. Further, in many teleosts, the natural mortality rate, M, approximates to 1.5 times the K parameter of the von Bertalanffy equation (Jensen 1996; Kenchington 2013b), while optimal fishing generally approximates to a fishing mortality rate equal to M. Thus, the optimum proportion of an exploited population that comprises individuals longer than $^2/_3\,L_\infty$ can be set as that seen when the total mortality rate, Z, is equal to $3K$ – as a first approximation pending development of species-specific values, though quantifying that proportion would need attention to the size-selective properties of the survey protocol. Those are, however, merely initial suggestions and their performances when applied to various data sets would need careful examination before their implementation in management contexts.

In summary, community-wide indices of size data have an attractive simplicity and might usefully serve to headline reports on environmental status but they require interpretation built on examination of both within-population trends in sizes and among-population shifts in the relative abundances of species with large and small adult sizes. Beneath each simplistic "headline", scientific advice should dwell on observed changes in units which are sufficiently finely divided that the causes of change can be identified and hence management responses, if any, can be selected. That need not always involve reporting on individual "stocks" but, in some cases, it might require examination of even more finely-divided sub-units – if inshore sub-populations were losing large fish, while offshore components of the same "stock" received strong recruitment, for example.

5.8 TROPHODYNAMICS

In contrast to indices of ecological diversity, which can be evaluated using field data on the relative abundances of species, ecosystem trophodynamics are not directly observable and hence trophodynamic metrics can only be founded on outputs of models of those ecosystems (ICES 2005) – albeit models that may themselves be based on field data. That extra and major step between ecosystems and measurements adds non-trivial additional uncertainty, requiring very careful attention to the veracity of the modelling, which has too often been lacking in past studies (cf. Rice 2000). Indeed, ICES (2005) concluded: "Many things have to improve before ecosystem tropho-dynamic models should be viewed as suitable sources for advice on specific management problems".

A wide variety of indices and indicators have nevertheless been suggested for tracking the trophic structures and functioning of marine ecosystems. Cury *et al.* (2005) were able to list more than two dozen generic classes of such measures relevant to the ecosystem approach to fisheries management. They selected six of those classes for further investigation, though the reasons for their choice were not reported. Of the half-dozen, quantifying the proportion of primary production which goes to support fish landings (termed "Primary Production Required" or "PPR") requires knowledge of consumption, ecotrophic efficiency and diet composition for each predator species – information that would seem to strain the limitations of existing scientific knowledge of most marine areas. The "Mixed Trophic Impact" measure requires an assumption of stability in trophic structure which has too obviously been violated in many aquatic ecosystems. Cury *et al.* (2005) made some use of ratios of biomasses of different trophic groups (see Section 5.2), as well as similar ratios of production and consumption of such groups – measures that may be useful as summaries of data, once observed trends have been successfully interpreted and understood. They gave perhaps the greatest prominence to "Mean Trophic Level", an index that has also been called the "Marine Trophic Index" ("MTI"), which was accompanied by the "Fishing-in-Balance" index of Pauly *et al.* (2000), both of which were developed in connection to Pauly *et al.*'s (1998a) observation of "fishing down marine food webs". Fishing-in-Balance scales overall catch (pooled across species) for the greater ecological efficiency of harvesting lower in the trophic pyramid, thus measuring one aspect of "fishing down" a trophic system. Its relevance to other issues is unclear. The MTI also has a role in quantifying Pauly *et al.*'s (1998a) observation but, in addition, has become one of the most widely used indices of marine biodiversity – despite its deficiencies for that purpose. Indeed, the MTI was the sole index amongst eight "indicators" selected in 2004 for "immediate testing" for CBD purposes[61]. It therefore provides a useful object lesson in the application of biodiversity metrics.

"Fishing down marine food webs" has come to mean two quite different things. In their formal paper, Pauly *et al.* (1998a) treated it as a straightforward observation of declining mean trophic level in fishery landings. That is now recognized as a slow but real change in many, though not all, of the world's fishing regions. When responding to subsequent criticisms by Caddy *et al.* (1998), Pauly *et al.* (1998b) emphasized that observational characteristic of their work. Much subsequent discussion has, however, treated "fishing down marine food webs" not as an emergent behaviour observed in a dataset but as a process – the transfer of fishing effort from high-trophic-level species to resources lower in the trophic pyramid consequent on the former being depleted by fishing. Thus, Pauly *et al.* (2000) explained "fishing down marine food webs" as "Large, slow-growing predators have been largely extirpated and replaced by small, fast-growing forage fish and invertebrates", while Bhathal and Pauly (2008) went on to write of "a process now widely known as 'fishing down marine food webs'" which is that "fisheries, upon depletion of the large, high-TL species they initially target, shift to small, low-TL species".

As designed, the MTI was simply a means for quantifying the observed decline in mean trophic level. That it can do, if not fully adequately, but it is not a measure of the hypothesized process driving the decline. The MTI has nevertheless been advanced as "an indicator of fishing impact on aquatic ecosystems" (Bhathal & Pauly 2008) and as a measure related to a management objective of "maintaining ecosystem structure and functioning" (Shin *et al.* 2010b), neither of which roles can it fulfil. BIP (2010) even described a prolonged decline in estimated MTI for selected fisheries in the defined Large Marine Ecosystems as "a global decrease in the abundance of high trophic level species", which the Index could never represent. Indeed, if it is anything more than a summary of questionable data, the MTI appears to be a poor measure of an aspect of human behaviour – specifically of one facet of fishermen's practices when targeting their fishing effort. Even if thus used, however, the MTI would confound changes in the trophic level of catches consequent on re-direction of effort in response to multiple drivers, not simply depletion of higher-level species but also management controls, technological development allowing economically viable seafood production from smaller (and hence likely lower trophic level) animals, and more (cf. Caddy *et al.* 1998; Essington *et al.* 2006).

---

[61] UNEP/CBD/COP/VII/30, February 2004.
The other seven "indicators" were: unspecified "trends", trends in abundances and distributions, coverage of protected areas, nitrogen deposition, water quality, linguistic diversity of indigenous languages, and developmental assistance provided in support of the Convention.

The MTI is simply a weighted average of the trophic levels of the species present in landings. As usually implemented (e.g. Cury *et al.* 2005; Pauly & Watson 2005; Bhathal & Pauly 2008; Shin *et al.* 2010b), each species is given a constant value for its trophic level, while the weighting is by the proportions of the species in fishery catches. The constancy of trophic level for each species sets aside the whole issue of the greater depletion of older (and hence typically larger and higher trophic level) individuals within a resource population which results from fishing mortality. The MTI thereby misses much of its point, leading Rogers and Greenaway (2005) to suggest that it would be better to focus on fish sizes and abundances than to adjust the available data using fixed multipliers. It must be said, however, that adequate data to describe size-specific diets are unavailable for most marine ecosystems. Even for such a long-studied case as the Atlantic cod of the western Scotian Shelf, Araújo and Bundy (2011) were only able to distinguish the diets of young-of-the-year and early juveniles (ages 1 to 3), all older and larger fish being lumped into a single dietary unit for lack of better data. For few other fish species of that area can subdivision by size or age proceed even that far (Cook & Bundy 2010; Araújo & Bundy 2011). Given such limited data, MTI cannot begin to respond to the spatial and temporal variability in diets that some fish show (Rogers & Greenaway 2005; ICES 2010). Meanwhile, the weighting by landings, rather than biomass, makes the MTI into a measure of anthropogenic pressure more than one of ecosystem biodiversity, though if it were intended as a pressure indicator, weighting by fishing mortality would be more appropriate.

Long after the MTI was adopted internationally, Branch *et al.* (2010) undertook what they described as “the first large-scale test of whether catch MTL is a good indicator of ecosystem MTL, marine biodiversity and ecosystem status”. That test might better have been conducted earlier, though the Index’s inadequacies in those roles should have been evident from inspection of its structure, coupled to a basic understanding of fisheries. Branch *et al.* (2010) found that the MTI does not track change in marine ecosystems and does not even track change in the mean trophic level of the resource populations, especially when fishing disproportionately affects one part of the food web – which was essentially the focus of Pauly *et al.*’s (2000) hypothesis explaining “fishing down marine foodwebs”. Branch *et al.* (2010) also confirmed what others had already reported: they found that global MTI had increased after the mid-1980s, in contradiction to Pauly *et al.*’s (1998a) observation of continuing decline through to 1995. The difference turned out to be a result of revisions to the estimates of species’ trophic levels in the “FishBase” on-line database, which had been the source used in both sets of calculations (Branch *et al.* 2010), though that application would seem to stretch the database’s capabilities. The reversal in MTI trends between the analyses of Pauly *et al.* (1998a) and those of Branch *et al.* (2010)

certainly casts doubt on any conclusions which might have been built on either result.

As seen in many another attempt to apply such broad metrics, Branch *et al.* (2010) resorted to extracting the information on particular fisheries from the data in order to understand their observed overall trends – which casts further doubt on the value of a summary metric. They found that a regime shift from low to high trophic levels in the Gulf of Alaska in the 1970s was hidden in the MTI by the high weighting given to the very large pollock fishery. Further, MTI increased in the Gulf of Thailand during the collapse of most resources there, since the sequence of fishery development did not follow the expectations of Pauly *et al.* (2000).

Faced with these severe deficiencies, there has been some attempt to apply the MTI calculations to regional biomass (*contra* global landings) estimates, with those derived either from survey time series or else from stock assessments. That approach allows for more careful estimation of species' trophic levels, since it can focus on far fewer species. Nicholson and Jennings (2004) applied that idea to data from the North Sea International Bottom Trawl Survey, using discrete trophic levels for each size class of each species (though ones fixed across time). They found, however, that the Index thus applied offered no greater statistical power than did their SBIs (see Section 5.7), while Piet and Jennings (2005) found that the MTI gave inconsistent responses. Thus, biomass-based MTI may be useful in long-term research studies but, at least in the North Sea, seems to have only limited value for supporting management decision-making. Branch *et al.* (2010) later examined multiple biomass-weighted applications of the Index, encountering difficulties with the biased selection of species for which biomass estimates are available, as well as the variable species' selectivities of survey gears. ICES (2010) also considered the biomass-weighted option but concluded that, even if so modified, the MTI would not be expected to show monotonic responses to anthropogenic pressures, some of which act on species of low trophic level (ICES 2010).

The final conclusions of Branch *et al.* (2010) merit quoting at length:

> Indicators such as catch MTL use readily available data and are quick and easy to calculate, but without improvement are ineffective measures of trends in biodiversity. [...] catch MTL does not reliably measure the magnitude of fishing impacts or the rate at which marine ecosystems are being altered by fishing. Instead, we recommend a greater emphasis on measuring and reporting changes in marine biodiversity by tracking trends in abundance [...].

That could stand as an epitaph for much of the effort devoted to monitoring the trophodynamic facet of biodiversity through the use of simplistic indices. All that need be added is that the deficiencies could be seen in the structure of the MTI from its first presentation and hence its failure was fully predictable.

Cury *et al.* (2005) applied variants of their six classes of trophodynamic indices to fishery data from the Benguela Current system during the second half of the 20th Century. They found that the indices tended towards stability, despite major changes having occurred in the fisheries. Cury *et al.* (2005) concluded that that stability "emphasizes the potential danger of interpreting a single indicator without analyzing the causes of the observed trajectory, or understanding the dynamics" of the system, which observation could be applied more widely than just to trophodynamics. Cury *et al.* (2005) went on to conclude: "Quantifying changes in an ecosystem is not straightforward, and no single trophodynamic indicator can track the complexity of the observed changes in fisheries and ecosystems" and suggested that "it might have been more informative to consider abundance and species composition from research surveys, rather than to compute the indicators used here".

Rather than summarizing the distribution of a community across trophic levels as a weighted mean (e.g. the MTI) or by another simple index, it is possible to represent that distribution as a "trophic spectrum" (Gascuel *et al.* 2005). Spectra might be drawn as biomass, abundance, extracted fishery catch or some other property of the community, against trophic level. That approach has great promise as a tool for exploring the structures of ecosystems and the changes in those structures under anthropogenic pressure, while the spectra themselves could be valuable illustrations of conclusions reached through deeper analysis of the condition of the ecosystem of interest. As with graphical methods for examination of species relative abundances, however, trophic spectra are not well suited to summary in simple, numerical terms and hence are poorly adapted for use in routine monitoring or mapping.

In short, trophodynamic indices may aid in reporting results and may serve to guide scientists towards issues meriting attention (though they may also guide attention away when either insensitivity or the conflicting consequences of contrasting trends in multiple drivers mask real change). What no simple index seems able to do is to reveal what is happening to an ecosystem. That needs broader and deeper analysis. While some authors have applied trophodynamic indices to dietary information derived directly from field studies (e.g. Bhathal and Pauly 2008), others have relied on the outputs from some model of the system, often one prepared using ECOPATH software (e.g. Cury *et al.* 2005). Once such models have been prepared, and supposing that they adequately represent reality (itself a major challenge), they can be used directly to examine changes across time, differences across space and the implications of each of those for management (e.g. Bundy 2005; Bundy & Fanning 2005; Araújo & Bundy 2012). There is then little value in reducing model outputs into any simple trophodynamic index.

5.9 DEMOGRAPHICS

Considering the vast amount of available information on the population dynamics of aquatic species, particularly but not exclusively the fishery resources, there has been curiously little attention to the potential of indices based on demographics as indicators of biodiversity (but see ICES 2005). Productivities are necessarily linked to mortality rates, which are in turn inversely related to life expectancies. The effects of fishery species-selection aside, fishing effort therefore tends to deplete long-lived, low-productivity species more than those naturally adapted to higher mortality rates. Meanwhile, within each resource species, the addition of fishing mortality on top of natural mortality shortens life expectancy and lowers mean and maximum ages in a population. Hence, indices based on observed ages, species maximum ages (in the absence of fishing), mortality rates or life expectancies could be applied in monitoring biodiversity. Aside from some investigation of mean lifespan and the biomass : landings ratio (as a proxy for the inverse of fishing mortality) by the *INDISEAS* group (cf. Shin *et al.* 2010b), we are not aware of any exploration of that possibility.

It may be wiser to leave the matter as it stands. For resource species, existing age compositions and total mortality rates (from which current life expectancies can be calculated) are generally well quantified but tracking those variables is a non-trivial task, requiring expensive data collection and analysis through stock assessments. It would be unwise to duplicate that work in support of biodiversity monitoring and impractical to extend it to non-resource species. Age compositions, mortality rates and life expectancies in the unfished condition are very much harder to estimate and the resulting estimates are necessarily surrounded by massive uncertainties (cf. Kenchington 2013b). While usefully-precise values are available for some species, any attempt to "score" the species in a community by such factors and then to generate a weighted-average of the species present would risk being grossly misleading if the uncertainties were ignored or entirely insensitive if they were considered.

5.10 CONSERVATION STATUS

The proportion of species in a community that have been formally recognized as "at risk" has been invoked as a biodiversity metric (cf. ICES 2010) and that may be the intent underlying the CBD's proposed monitoring of extinction risks[62]. Clearly, monitoring the status of particular species that may be at risk is a most important aspect of information collection in support of biodiversity conservation. Unfortunately, the abundances of such species are not often adequately quantified by general surveys (whether designed for fishery-resource biomass estimation or community-diversity measurement), if only because extreme rarity makes the capture of specimens too uncertain but often also because the at-risk species are atypical – sea turtles or baleen whales, rather than groundfish, for example (cf. ICES 2012b). Hence, species-specific monitoring programs are usually required. By definition, the data from those lie outside the scope of biodiversity metrics, even though needed in support of biodiversity conservation. When the focus shifts from monitoring species at risk to using the number of species at risk as an index of some facet of biodiversity, however, difficulties arise (Quayle & Ramsay 2005; ICES 2012b). The number of species formally listed is too largely determined by progress with the bureaucratic (and highly politicized) task of listing, a process which also tends to be biased towards a focus on charismatic megafauna. Its responses to the actual conservation status of the species concerned is likely to be much weaker. Even if the listing process worked perfectly, it is based on scientific perceptions of conservation and taxonomic statuses, which develop separately from the statuses themselves, such that a species newly listed may actually have been at risk for a prolonged period. Indeed, some naturally-rare species may have met modern listing criteria even in historic times. Besides, the most common way that species leave an "at risk" list is through extinction or extirpation. Thus, a decline in the proportion of species officially listed is likely to indicate an increase in environmental degradation, whereas increases signify progress with paperwork more than worsening degradation. Quayle and Ramsay (2005) thus described the use of a threatened-species list in tracking temporal change in biodiversity as "inappropriate and potentially misleading", urging instead long-term monitoring of such real biological features as abundances and range sizes.

---

[62] UNEP/CBD/COP/DEC/XI/3, December 2012.

The proportion of fishery-resource "stocks" that are currently fished at or below their optimum exploitation rates, *versus* those overfished, has similar appeal to metrics summarizing "at-risk" status (cf. Shin *et al.* 2010b; ICES 2012a) but bitter experience has shown that existing stock assessments are often too imprecise for the condition of the resources to be judged with useful precision until long after the event – the rapid transition of the perceived status of the NAFO Divisions 2J3KL cod from sustainably fished in the later 1980s to grossly over-exploited in the early 1990s being merely the most infamous example. Fisheries management requires the support of timely, if highly uncertain, assessments. Monitoring and mapping of biodiversity are not so time-sensitive and hence need not engage with the uncertainties resulting from the pressures on stock assessment.

In a similar vein in 2008, the Commission of the European Community issued a Decision adopting a program for the collection, management and use of data concerning the fisheries sector. That provided nine indicators, four of which could be considered as measures of biodiversity. They included three Size-Based Indicators plus "Conservation Status of Fish Species". The latter requires a list of fish species (for a particular survey series) selected as both "sensitive to fishing", though that may only mean that the largest individuals of each species exceed a set length, and abundant enough that more than 20 individuals are taken in resource surveys each year, while species with low and variable catchability in the survey gear and those which cannot be reliably identified are excluded. That list is then reduced to just 20 species, if the initially selected number is greater, by discarding all but the largest – "large" being measured as each population's asymptotic length. From the survey data and for each selected species, an abundance index is then determined for individuals longer than one half the length of the 95th percentile of population lengths (apparently intended as a proxy for length-at-maturity) and the ten-year trend in those abundance values is assessed against IUCN criteria for endangerment, scored as 3 for "critically endangered", 2 for "endangered", 1 for "vulnerable" and otherwise zero (that step echoing Dulvy *et al.* 2006). The average of those species scores, across all 20 listed species, comprises the "Conservation Status" index. Subsequent to its adoption, that metric was referred to ICES for comment, which led to its sensitivity being questioned (ICES 2012a). That the index averages the statuses of up to 20 species (rather than focusing on those declining fastest) was seen as potentially masking problems, while invoking the IUCN criteria imposes a linearity on temporal trends in abundance which can be unrealistic. Moreover, the process of determining index values proved to be too subjective. ICES (2012a) indicated a preference for a simpler and more easily interpretable way of describing trends common to multiple species. The EC "Conservation Status" measure certainly seems too ill-considered an index to merit being imported into Canada.

5.11 INTRA-SPECIFIC DIVERSITY

In Section 2.4, we stressed the importance, in aquatic ecosystems and especially in the open sea, of declines in intra-specific biodiversity, including the loss of genetic diversity. Maintaining the structure of local populations within a species is a key step in genetic conservation, with special attention being given to source populations that supply recruitment to areas downstream. Unfortunately, direct monitoring of allele frequencies remains too expensive for routine application to most species, Where the cost is deemed appropriate, as it may be for some species at risk, the best index for representing allelic diversity is Jost's (2008) Δ, as explained in Section 3.3 above, though Meirmans and Hedrick's (2011) caution about the use of Jost's (2008) metrics with multi-locus data should not be ignored. Whether the more suitable option is $\Delta_{(q=1)}$ or $\Delta_{(q=2)}$, as well as the best way to use the indices with multi-locus data, remain to be determined by population geneticists.

Should a program for monitoring allele frequencies be designed, it would be essential to clearly define the management objective to be supported by the monitoring. The scope for various kinds of genetic monitoring is wide, while different objectives raise different questions and hence different monitoring to provide the answers. Hence, appropriate genetic markers and monitoring methods cannot be selected until the objectives have been established (Kenchington & Heino 2002; Kenchington *et al.* 2003).

More often, routine monitoring of intra-specific diversity will track population parameters and will be linked to genetics through research work, which lies outside the scope of the present report. In practice, the most important management goal is probably maintenance of population structure, ensuring an adequate abundance of each of the local units (cf. ICES 2005). That requires an initial description of the structure, including but not limited to documentation of allele frequencies in the various populations. Thereafter, population-level monitoring of numbers and biomasses, perhaps with size and age data also, should suffice. Checking that the biomass in a local area continues to be composed of individuals having the phenotypic and genotypic characteristics expected of the resident population would only be needed at rare intervals. While that approach poses a substantial requirement for baseline research and demands more intensive monitoring than is provided by most existing survey series (which only seek to quantify the biomasses in each "stock" rather than those in each sub-unit), there is no need for any novel indices.

Likewise, intermittent research studies aside, "Fisheries-Induced Evolution" can be tracked by monitoring sizes-at-age, sizes at first maturity and other demographic measures. For resource species, most of the required data are already routinely gathered and have been for many years. Some expansion of monitoring to cover those bycatch species which may be vulnerable to FIE may be appropriate but would not involve any special challenges.

Where even population-level monitoring of biomass is deemed unreasonably onerous, monitoring of broader spatial distributions and environmental conditions may have some value, though it cannot fully substitute for data on individual populations. Maintenance of the full geographic range of each species would help ensure that unrecognized population structure is preserved, while continuation of natural environmental conditions throughout that range would provide stability in selective factors and hence preserve allele frequencies (cf. Kenchington & Heino 2002; Kenchington *et al.* 2003). Perfect maintenance of natural ranges and conditions would, however, be inconsistent with natural changes in ocean climates, with the elevated mortality rates and reduced abundances that are inevitable consequences of fishing, and with the on-going consequences of other human uses of aquatic systems. Management requires optimal balances between use and preservation, not maximization of either, and that requirement drives a need for scientific advice based on knowledge built on extensive monitoring data. Exclusively species-level monitoring may assist but can never be sufficient.

5.12 SUMMARY

There is no doubt of the importance of many facets of biodiversity, to human interests as well as to the intrinsic values of natural systems. Community structures (in the sense of relative abundances of species and other taxa), size compositions, functional traits, trophic structures, diversity of habitats and particularly intra-specific diversity are all critical concerns. The metrics that have been advanced for the quantification of those facets are not without merit but they are neither efficient nor effective.

As argued in previous Sections of this report, the metrics of ecological diversity are poor measures of community structure. They serve as overly-simplistic summaries of Species Abundance Distributions, and hence of community structure, but they are essentially devoid of higher meaning, while they fail to capture much of the information in the SADs. Since the latter are themselves distorted images of real biological communities, being constrained by spatial and temporal bounds, then filtered through survey protocols, quantitative values for ecological diversity may be needed but they have limited utility. The only difference between those metrics and indices of other facets of biodiversity is that those other facets can be more readily perceived as distinct from the means used in their measurement: ecological diversity has always been conceived in terms of simple indices but trophic structures, for example, have not. We have shown in this Section that attempts to measure the other facets using indices are as ineffective as the attempt to capture community composition in metrics of ecological diversity. That is true in general but also for those particular indices that have been widely promoted as broadly meaningful measures of biodiversity and of anthropogenic impacts, rather than merely as crude summaries of data. Closer examination shows that Δ^+, Λ^+, the LPI, the LFI and the MTI at best carry no more meaning than does H' and, at worst, can be grossly misleading. Mayer (2006) saw that biodiversity "cannot be caught in numbers" but thought that single facets could be measured. We do not disagree but it seems that even a single facet cannot usefully be caught by any simple metric.

That conclusion runs counter to the widespread expectation that measurement of biodiversity should use simple, globally-applicable indices – the sort of metrics that are the focus on the present report. In the era before abundant, available computing power, the attempt to represent ecological diversity by a single scalar value was appropriate but, as a direct consequence of that restriction, the resulting measurements had little ecological meaning or conservation relevance. Research into quantitative ecology therefore moved away from diversity measures 30 years ago. The renewed interest in biodiversity, driven by advocacy more than science, has led to a major broadening of policy responses to anthropogenic pressures on ecosystems. It would be a serious mistake to discard that advance by falling back onto narrow considerations simply because those are easily represented in numbers. Yet simple indices must always miss the point of biodiversity. Distilling the almost-unlimited complexity of a real ecosystem, even a single facet of a real ecosystem, into a scalar value requires that almost all the complexity be discarded, and with it most of the information in an expensively collected dataset. It is, however, the complexities of ecological systems that comprise biodiversity and stripping those away in order to quantify them must always be dysfunctional.

The pressure to seek simple metrics is understandable, nonetheless. The complications of real ecosystems are so intricate as to overwhelm both human comprehension and current scientific descriptive capabilities. Faced with that challenge, the basic training of scientists encourages a search for abstractions and simplifications that cut through the confusion and, by intention, reveal underlying patterns – precisely the motivation that led to the concepts of ecological diversity half a century ago. Too many have succumbed to that temptation, creating a precipitate rush into development of ever more metrics of doubtful utility. Meanwhile, some management decision-makers, not to mention many journalists and much of the lay public, seek to have scientific understanding of complex issues boiled down to simple measures that at least appear understandable, preferably measures in the form of one-dimensional scores or even a lone binary, "good" *vs.* "bad", indicator. Scientific advisors should comprehend the complexities and resist the over-simplification. A few have done so and there has been a persistent drum-beat of cautionary warnings, which have recently been summarized by Green and Chapman (2011). Their bold conclusions are best stated in their own words:

> scientists should not be developing or relying on single number representations of complex data. And they should not be misleading non-scientists that this is appropriate or even useful. [...]
> At best reduction to an index means loss of information. [...] indices can be misleading and, if used at all, should not be used in isolation. [...]
> there is a common attitude that the implementation of [multivariate] approaches and assessment of their output are too complex to transmit easily to managers [...]. Perhaps what we need is better managers and better education of environmental scientists; [...]
> If indices must be calculated and presented then this should be done together with other statistical methods that retain more of the information in the biological data set [...]
> avoid using indices because of information loss and the likelihood that their use will lead to misleading conclusions. If you absolutely must use indices for some non-scientific reason [...], use them together with other statistical methods that retain more of the information in the biological data set. Developing simplistic numbers simply to satisfy the least knowledgeable scientists and managers is hardly the best way to advance either scientific knowledge or management decision-making.

Our examination of the available indices of biodiversity leads us to a wholehearted agreement with Green and Chapman's (2011) position. Bevilacqua *et al.* (2012a) reached a rather similar conclusion. They noted Pielou's (1975) remark that "The belief (or superstition) of some ecologists that a diversity index provides a basis (or talisman) for reaching a full understanding of community structure is wholly unfounded" and themselves declared that the search for "a single metric able to serve as a one-fit-all solution for environmental issues and easily [*sic*] to communicate to environmental managers" is a "quest [..] probably doomed to failure, as any endeavor to condense biological and ecological complexity into a single number". Their alternative, which followed Green and Chapman (2011) was particularly telling: "better managers and better education of environmental scientists".

Contrary suggestions have been published but have not been given much rational foundation. Pauly and Watson (2005) declared "to be of any use, indicators must summarize in a single number a variety of complex processes that are otherwise hard to apprehend", a notion that Bhathal and Pauly (2008) rendered as "a need for predictive indicators that can be easily parameterized using easily accessible data, while communicating with a single number a variety of complex processes occurring with an ecosystem". Yet neither they nor the authors that they cited in support offered any argument for why it might be helpful for scientists to summarize multiple processes in a single indicator, still less a single number, which seems to us a most unhelpful approach. In an examination of marine ecosystem indicators that made little mention of biodiversity but addressed what is here termed environmental status, Rogers and Greenaway (2005) argued that rather simple measures are useful when applied to single issues and especially to particular anthropogenic pressures. For nutrient enrichment and the risk of eutrophication, however, they warned that: "It is essential to be able to complete this chain of cause and effect if anthropogenic and natural environmental change are to be distinguished and these indicators are ever to be used for management". Where cause-and-effect relationships can be inferred, simple indicators are indeed appropriate as a foundation for decision-making. That requirement can never be met, however, for high-level summaries of ecosystems such as biodiversity measures. Greenstreet and Rogers (2006) made the valid point that: "If the conservation and management of biodiversity are management objectives [...] then indicators of biodiversity will be required" but they did not attempt to justify their decision to confine those indicators to ones quantified through simple indices. More succinctly, Borja *et al.* (2008) stated the goal of indices as "to reduce or summarize environmental conditions or quality to a number, which will form the basis for management decisions", while Borja and Dauer (2008) suggested that the aim is to summarize and simplify complex data, as well as to convey information in a manner easily understood by non-scientists. Yet that is but to encourage decision-makers to focus on over-simplifications, when the value of biodiversity is wrapped up in its complexity. In the context of indicators of the state of exploited marine ecosystems, Shin *et al.* (2010a) posed the question "Can simple be useful and reliable?" but they did not answer it, instead proceeding from a set of simple indices to an even simpler presentation of one pie diagram per broad marine region.

Degnbol (2005) has provided a more thoughtful rationale for the use of indicators, though his starting point was the need for a partnership of institutions and stakeholders in addressing cognitive and normative aspects of management, which requires that indicators be evaluated not only technically but also for their effectiveness in communicating knowledge. We agree. No indicator, however, can communicate knowledge unless it captures meaningful information and does not conceal it under an appearance of meaning something else. Indices designed to reduce the complexities of ecosystems to a few numbers cannot achieve that standard. Notably, Degnbol (2005) concluded that indicators which satisfy both stakeholders and scientists "tend to be system-specific, and may not be compared across systems". We suggest that that tendency can also be seen in indicators truly satisfactory for either group, not only those that meet joint requirements. Similarly, Johnson (2008) declared that "environmental indicators have a crucial role to play in the simplification, quantification, standardization, and rational explanation (communication) of environmental conditions to regulators and policy-makers" but he prefaced that with: "Assuming that particular metrics tell us much about either rates or extent of change ...".

We have previously suggested that the indices of ecological diversity must be calculated because of the broad expectation that they will be. Those should be accompanied by measures of overall abundance of the taxocene in question (see Section 5.2). They could be supplemented with a metric of phylogenetic diversity, if a suitable tree is available. That would capture more information that any index of ecological diversity can but not much more, while phylogenetic measures introduce complexity and risk inducing confusion through lack of general familiarity. Allelic diversity can certainly be quantified, if the data are available, though Whitlock (2011) might argue that G_{ST}, and the other F_{ST} analogues are more important, along with the demographic inferences that can be built from them. Otherwise, what is required is not an array of simplistic metrics but rather analytical tools that can capture the information in datasets and reveal the patterns in that information. Most of those tools likely involve multivariate statistics but, we suggest, the important issue is not methodological but rather one of efficient and effective extraction of knowledge about complex ecosystems from available data. No simple metric can provide that.

6 TOWARDS QUANTIFICATION OF BIODIVERSITY

If biodiversity, and indeed environmental status, can only be quantified facet by facet (cf. Mayer 2006; Sarkar 2010), indeed often only as that facet is reflected by a particular taxocene (as filtered through some survey protocol), and if few (or none) of those facets can usefully be quantified using simple metrics (see Section 5.12), a question must arise of how biodiversity can be measured – as it must be if rational conservation policies are to be implemented. We propose that the measurement should employ approaches which can capture the depth and complexity of biodiversity. They should be more powerful and more flexible, and hence inevitably far more complex, than the simple indices reviewed in this report. Indeed, the need is not for a single type of analysis, such as some ordination, but rather for "biodiversity assessments", analogous to the stock assessments prepared in support of fisheries management and, like those, utilizing a suite of analytical tools adapted to case-specific requirements. The details must depend on the facet of biodiversity being measured, the available datasets and particularly the patterns of diversity captured in the data and revealed by the analyses. Multivariate statistical approaches would likely be involved but there would be few broadly-applicable generalities. While ecologists, population genetics and fisheries scientists have decades of experience with different arrays of appropriate techniques, little of that work has been conceived as biodiversity measurement. Thus, there is a need for an evolutionary period of methodological development, during which scientists engaged in biodiversity assessment should experiment with different combinations of analyses to determine which are the most informative or particular cases – a process begun by ICES (2012b), if not before.

This recognition of a need for biodiversity assessments may be novel but the requirement should not be surprising. When seeking scientific support for groundfish fishery management, the Department has long dispatched research trawlers to sea, where they have conducted routine surveys that gather large amounts of data. For any one survey and any one "stock", those data could be boiled down to an annual scalar value, such as the mean catch per standard set. More often, the survey data are summarized as arrays of numbers-at-age and weights-at-age across time. Such an array may be included amongst the scientific advice to fisheries managers, perhaps in the form of a figure in a *Stock Status Report*, but the array is not itself the scientific advice. Rather, it is combined with many other sources of information, fed into a well-founded mathematical model of the resource population, the output subjected to professional examination and interpretation, followed by peer review, from which the advice emerges. If that is the degree of complexity necessary when examining the status of one population and its responses to one driver, fishing effort, it should not be expected that the data from the same survey series can be boiled down to a time series of scalar values that could provide useful indications of the status of an entire ecosystem, subject to multiple major drivers. If fishery resources need stock assessments, as they clearly do, then ecosystems need biodiversity assessments even more.

That is not to say that simple indices have no role. Scientific capacity to measure the state of an ecosystem far surpasses the ability of single humans, even highly trained scientists, to comprehend the mass of data. Decision-makers not merely desire but need scientific advice that cuts through the complexity. The questions are how far to simplify in presenting scientific advice to end-users and where, within the advisory process, to undertake that simplification. From our review of the available metrics, we conclude that they are not suited to being the means by which scientists develop their advice but they do have a place in communicating that advice – provided that they do not substitute for it. Any presentation of a "state of the ecosystem" report should probably include a summary statement, which will necessarily come close to being one-dimensional, maybe even binary. Many assessments would lead to conclusions, derived from advanced analyses, which can be illustrated by one or more time series or maps of the values of some simple index. The difference from past practice should be that the index used in communicating advice would be selected after the conclusions have been reached and would not be the foundation of them. Potentially, if some "benchmark" biodiversity assessment identified a time series of a particular metric as illustrative of trends in the condition of an ecosystem, then extensions of that series could serve to update understanding of the system, pending the next full assessment. That, however, is as far as simple indices should be trusted as measures of biodiversity.

Downgrading indices to little more than vehicles for communicating conclusions, essentially headlines for paragraphs of advice, carries a risk that end-users, conditioned to expect indices and lacking time for further reading, will focus on simple numbers and thus be misled. Great care would therefore be needed in the presentation of index values. Rice (2003) considered some of the challenges involved in using indicators in communicating results and his comments merit attention by those charged with the task. He did not, however, explore the common situation in which the recipients of the advice suppose that they understand the message conveyed by the indicator, when they in fact misunderstand it – a challenge that communicators of advice must never forget.

The proposed assessments would not be cheap. Biodiversity is complex. Measuring it, in any meaningful way, must be expensive. While cost will always be an important consideration, there is no value in spending on measurements that are meaningless. If meaningful measurement, based on a biodiversity assessment, is beyond the scope of available budgets, then an absence of measurement should be accepted, not hidden with meaningless metrics. Besides the cost, abandoning over-simplified indices would mean not using those few measures of biodiversity that already enjoy some public awareness and not using any substitute that is "concrete", in Rice and Rochet's (2005) sense, and hence easily explained. At least, the analytical techniques used in biomass assessment are likely to be as opaque to end-users as the virtual population analyses of stock assessments have long been. However, where the results of those assessments lend themselves to illustration using simple indices, familiar and "concrete" measures could be used to communicate conclusions.

ACKNOWLEDGEMENTS

This work was supported by funding under the Department's Strategic Program for Ecosystem-Based Research and Advice (SPERA) to ELK. We thank Dr. Claudio DiBacco and F. Javier Murillo for very helpful comments on an earlier draft but take sole responsibility for the conclusions and recommendations offered herein.

REFERENCES

Abellán, P., Bilton, D.T., Millán, A., Sánchez-Fernández, D., and Ramsay, P.M. 2006. Can taxonomic distinctiveness assess anthropogenic impacts in inland waters? A case study from a Mediterranean river basin. Freshwater Biology 51: 1744-1756.

Ames, E.P. 2004. Atlantic cod stock structure in the Gulf of Maine. Fisheries, Bethesda 29(1): 10-28.

Anderson, P.K. 1995. Competition, predation, and the evolution and extinction of Steller's sea cow, *Hydrodamalis gigas*. Marine Mammal Science 11: 391-394.

Anon. 2012. UK Biodiversity Indicators in Your Pocket. Department for Environment, Food and Rural Affairs, London: 37 p.

Araújo, J.N., and Bundy, A. 2011. Description of three Ecopath with Ecosim ecosystem models developed for the Bay of Fundy, western Scotian Shelf and NAFO Division 4X. Can. Tech. Rep. Fish. Aquat. Sci. 2952: xii+189 p.

Araújo, J.N., and Bundy, A. 2012. Effects of environmental change, fisheries and trophodynamics on the ecosystem of the western Scotian Shelf, Canada. Mar. Ecol. Prog. Ser. 464: 51-67.

Backer, H. 2008. Indicators and scientific knowledge in regional Baltic Sea environmental policy. ICES J. Mar. Sci. 65: 1398-1401.

Balvanera, P., Pfisterer, A.B., Buchmann, N., He, J.-S., Nakashizuka, T., Raffaelli, D., and Schmid, B. 2006. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. Ecology Letters 9: 1146-1156.

Beaumont, N.J., Austen, M.C., Atkins, J.P., Burdon, D., Degraer, S., Dentinho, T.P., Derous, S., Holm, P., Horton, T., van Ierland, E., Narboe, A.H., Starkey, D.J., Townsend, M., and Zarzycki, T. 2007. Identification, definition and quantification of goods and services provided by marine biodiversity: Implications for the ecosystem approach. Mar. Poll. Bull. 54: 253-265.

Beisel, J.-N., Usseglio-Polatera, P., Bachmann, V., and Moreteau, J.-C. 2003. A comparative analysis of evenness index sensitivity. Internat. Rev. Hydrobiol. 88: 3-15.

Benoît, H.P., and Swain, D.P. 2003. Standardizing the southern Gulf of St. Lawrence bottom-trawl survey time series: Adjusting for changes in research vessel, gear and survey protocol. Can. Tech. Rep. Fish. Aquat. Sci. 2505: iv+95 p.

Bevilacqua, S., Fraschetti, S., Musco, L., Guarniere, G., and Terlizzi, A. 2011. Low sensitiveness of taxonomic distinctness indices to human impacts:

Evidences across marine benthic organisms and habitat types. Ecological Indicators 11: 448-455.

Bevilacqua, S., Sandulli, R., Plicanti, A., and Terlizzi, A. 2012a. Taxonomic distinctness in Mediterranean marine nematodes and its relevance for environmental impact assessment. Mar. Poll. Bull. 64: 1409-1416.

Bevilacqua, S., Terlizzi, A., Claudet, J., Fraschetti, S., and Boero, F. 2012b. Taxonomic relatedness does not matter for species surrogacy in the assessment of community responses to environmental drivers. J. Appl. Ecol. 49: 357-366.

Bhathal, B., and Pauly, D. 2008. 'Fishing down marine food webs' and spatial expansion of coastal fisheries in India, 1950–2000. Fisheries Research 91: 26-34.

Biodiversity Indicators Partnership (BIP) 2010. Biodiversity Indicators and the 2010 Target: Experiences and lessons learnt from the 2010 Biodiversity Indicators Partnership. Secretariat of the Convention on Biological Diversity, Montréal, Canada, Technical Series 53: 196 p.

Biodiversity Indicators Partnership (BIP) 2011. Guidance for National Biodiversity Indicator Development and Use. UNEP–World Conservation Monitoring Centre, Cambridge: 40 pp.

Bird, C.E., Karl, S.A., Smouse, P.E., and Toonen, R.J. 2011. Detecting and measuring genetic differentiation. *In* Crustacean Issues: Phylogeography and Population Genetics in Crustacea. Edited by S. Koenemann, C. Held and C. Schubart. CRC Press, Boca Raton, Florida. pp. 31–55.

Borja, A., and Dauer, D. 2008. Assessing the environmental quality status in estuarine and coastal systems: comparing methodologies and indices. Ecological Indicators 8: 331-337.

Borja, A., Dauer, D., Díaz, R., Llansó, R.J., Muxika, I., Rodríguez, J.G., and Schaffner, L. 2008. Assessing estuarine benthic quality conditions in Chesapeake Bay: A comparison of three indices. Ecological Indicators 8: 395-403.

Borja, A., Franco, J., and Pérez, V. 2000. A Marine Biotic Index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Mar. Poll. Bull. 40: 1100-1114.

Borja, A., Muxika, I., and Franco, J. 2003. The application of a Marine Biotic Index to different impact sources affecting soft-bottom benthic communities along European coasts. Mar. Poll. Bull. 46: 835-845.

Borja, A., Ranasinghe, A., and Weisberg, S.B. 2009. Assessing ecological integrity in marine waters, using multiple indices and ecosystem components: Challenges for the future. Mar. Poll. Bull. 59: 1-4.

Borja, A., and Tunberg, B.G. 2011. Assessing benthic health in stressed subtropical estuaries, eastern Florida, USA using AMBI and M-AMBI. Ecological Indicators 11: 295-303.

Branch, T.A., Watson, R., Fulton, E.A., Jennings, S., McGilliard, C.R., Pablico, G.T., Ricard, D., and Tracey, S.R. 2010. The trophic fingerprint of marine fisheries. Nature 468: 431-435.

Buch, E., Pedersen, S.A., and Ribergaard, M.H. 2004. Ecosystem variability in West Greenland waters. J. Northw. Atl. Fish. Sci. 34: 13-28.

Buckland, S.T., Magurran, A.E., Green, R.E., and Fewster, R.M. 2005. Monitoring change in biodiversity through compound indices. Phil. Trans. Roy. Soc. B 360: 243-254.

Buckland, S.T., Studeny, A.C., Magurran, A.E., Illian, J.B., and Newson, S.E. 2011a. The geometric mean of relative abundance indices: a biodiversity measure with a difference. Ecosphere 2: art100. doi:10.1890/ES11-00186.1

Buckland, S.T., Studeny, A.C., Magurran, A.E., and Newson, S.E. 2011b. Biodiversity monitoring: the relevance of detectability. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 25-36.

Bundy, A. 2005 Structure and functioning of the eastern Scotian Shelf ecosystem before and after the collapse of groundfish stocks in the early 1990s. Can. J. Fish. Aquat. Sci. 62: 1453-1473.

Bundy, A., and Fanning, L.P. 2005. Can Atlantic cod (*Gadus morhua*) recover? Exploring trophic explanations for the non-recovery of the cod stock on the eastern Scotian Shelf, Canada. Can. J. Fish. Aquat. Sci. 62: 1474-1489.

Byrnes, J.E., Reynolds, P.L., and Stachowicz, J.J. 2007. Invasions and extinctiuons reshape coastal marine food webs. PLos ONE 2(3): e295.

Caddy, J.F., Csirke, J., Garcia, S.M., and Grainger, R.J.R. 1998. How pervasive is "fishing down marine food webs"? Science 282: 183.

Callaway, R., Alsvåg, J., de Boois, I., Cotter, J., Ford, A., Hinz, H., Jennings, S., Kröncke, I., Lancaster, J., Piet, G., Prince, P., and Ehrich, S. 2002. Diversity and community structure of epibenthic invertebrates and fish in the North Sea. ICES J. Mar. Sci. 59: 1199-1214.

Callaway, R., Engelhard, G.H., Dann, J., Cotter, J., and Rumohr, H. 2007. A century of North Sea epibenthos and trawling: comparison between 1902-1912, 1982-1985 and 2000. Mar. Ecol. Prog. Ser. 346: 27-43.

Campana, S., Fanning, P., Fowler, M., Frank, K., Halliday, R., Lambert, T., Mohn, R., Wilson, S., and Stobo, W. 1995. Report of the 4Vn cod working group on the scientific value of a 4Vn cod (May-Oct) stock assessment. DFO Atl. Fish. Res. Doc. 95/16.

Cardinale, M., Svedang, H., Bartolino, V., Maiorano, L., Casini, M., and Linderholm, H. 2012. Spatial and temporal depletion of haddock and Pollack during the last century in the Kattegat-Skagerrak. J. Applied Ichthyol. 28: 200-208.

Chao, A., Chiu, C.-H., and Jost, L. 2010. Phylogenetic diversity measures based on Hill numbers. Phil. Trans. Roy. Soc. B 365: 3599-3609.

Chao, A., Colwell, R.K., Lin, C.-W., and Gotelli, N.J. 2009. Sufficient sampling for asymptotic minimum species richness estimators. Ecology 90: 1125-1133.

Chao, A. and Shen, T.-J. 2003. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. Environ. Ecol. Statistics 10: 429-443.

Charnov, E.L., Gislason, H., and Pope, J.G. 2012. Evolutionary assembly rules for fish life histories. Fish and Fisheries doi 10.1111/j.1467-2979.2012.00467.x

Chown, S.L., and McGeoch, M.A. 2011. Measuring biodiversity in managed landscapes. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 252-264.

Clarke, K.R., and Warwick, R.M. 1998. A taxonomic distinctness index and its statistical properties. J. App. Ecology 35: 523-531.

Clarke, K.R., and Warwick, R.M. 2001. A further biodiversity index applicable to species lists: variation in taxonomic distinctness. Mar. Ecol. Prog. Ser. 216: 265-278.

Cook, A.M., and Bundy, A. 2010. The food habits database: an update, determination of sampling adequacy and estimation of diet for key species. Can. Tech. Rep. Fish. Aquat. Sci. 2884: iv+144 p.

Crawford, N.G. 2010. SMOGD: software for the measurement of genetic diversity. Molecular Ecology Resources 10: 556-557.

Cury, P.M., Shannon, L.J., Roux, J.-P., Daskalov, G.M., Jarre, A., Moloney, C.L., and Pauly, D. 2005. Trophodynamic indicators for an ecosystem approach to fisheries. ICES J. Mar. Sci. 62: 430-442.

Dauvin, J.-C., and Ruellet, T. 2009. The estuarine quality paradox: Is it possible to define an ecological quality status for specific modified and naturally stressed estuarine ecosystems. Mar. Poll. Bull. 59: 38-47.

Degnbol, P. 2005. Indicators as a means of communicating knowledge. ICES J. Mar. Sci. 62: 606-611.

Department of Fisheries and Oceans (DFO) 2010. Review of the Manicouagan Marine Protected Area (MPA) ecological monitoring plan. DFO CSAS Sci. Adv. Rep. 2010/075.

Department of Fisheries and Oceans (DFO) 2011. Musquash Estuary: a proposed monitoring framework for the Marine Protected Area and intertidal area administered by Fisheries and Oceans Canada. DFO CSAS Sci. Adv. Rep. 2011/040.

DeVaney, S.C., Hartel, K.E., and Themelis, D.E. 2009. The first records of *Neocyema* (Teleostei: Saccopharyngiformes) in the western North Atlantic with comments on its relationship to *Leptocephalus holti* Schmidt 1909. Northeastern Naturalist 16: 409-414.

Diaz, R.J., Solan, M., and Valente, R.M. 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. J. Environ. Manage. 73: 165-181.

Domning, D.P., Thomason, J., and Corbett, D.G. 2007. Steller's sea cow in the Aleutian Islands. Mar. Mamm. Sci. 23: 976-983.

Dornelas, M., Magurran, A.E., Buckland, S.T., Chao, A., Chazdon, R.L., Colwell, R.K., Curtis, T., Gaston, K.J., Gotelli, N.J., Kosnik, M.A., McGill, B., McCune, J.L., Morlon, H., Mumby, P.J., Øvreås, L., Studeny, A., and Vellend, M. 2012. Quantifying temporal change in biodiversity: challenges and opportunities. Proc. Roy. Soc. B 280: 20121931.

Dornelas, M., Soykan, C.U., and Ugland, K.I. 2011. Biodiversity and disturbance. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 237-251.

Duffy, J.E. 2009. Why biodiversity is important to the functioning of real-world ecosystems. Front. Ecol. Environ. 7: 437-444.

Dulvy, N.K., Jennings, S., Rogers, S.I., and Maxwell, D.L. 2006. Threat and decline in fishes: an indicator of marine biodiversity. Can. J. Fish. Aquat. Sci. 63: 1267-1275.

Dulvy, N.K., Pinnegar, J.K., and Reynolds, J.D. 2009. Holocene extinctions in the sea. *In* Holocene Extinctions. Edited by S.T. Turvey. Oxford U.P., Oxford. pp. 129-150.

Dulvy, N.K., Sadovy, Y., and Reynolds, J.D. 2003. Extinction vulnerability in marine populations. Fish and Fisheries 4: 25-64.

Edelaar, P., Burraco, P., and Gomez-Mestre, I. 2011. Comparisons between Q_{ST} and F_{ST} — how wrong have we been? Molecular Ecology 20: 4830-4839.

Ellis, J.R., Maxwell, T., Schratzberger, M., and Rogers, S.I. 2011. The benthos and fish of offshore sandbank habitats in the southern North Sea. J. Mar. Biol. Asso. UK 91: 1319-1335.

Ellison, A.M. 2010. Partitioning diversity. Ecology 91: 1962-1963.

Essington, T.E., Beaudreau, A.H., and Wiedenmann, J. 2006. Fishing through marine food webs. PNAS 103: 3171-3175.

European Commission (EC) 2003. Common Implementation Strategy for the Water Framework Directive (2000/60/EC): Guidance Document n°. 5: Transitional and Coastal Waters – Typology, Reference Conditions and Classification Systems. Office for Official Publications of the European Communities, Luxembourg: vi+107 p.

Fanning, L.P. 1985. Intercalibration of research survey results obtained by different vessels. CSAS Res. Doc. 1985/003.

Fisher, R.A., Corbet, A.S., and Williams, C.B. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. J. Animal Ecol. 12: 42-58.

Fogarty, M.J., and Murawski, S.A. 1998. Large-scale disturbance and the structure of marine systems: fishery impacts on Georges Bank. Ecological Applications 8: S6-S22.

Fowler, G.M. 2011. Old and older perceptions of the migrations and distribution of haddock, *Melanogrammus aeglefinus*, in northwest Atlantic waters from tagging conducted in the Bay of Fundy, Georges Bank, Scotian Shelf, and the southern Gulf of St. Lawrence. J. Northw. Atl. Fish. Sci. 43: 137-157.

Fowler, G.M., and Showell, M.A. 2009. Calibration of bottom trawl survey vessels: Comparative fishing between the Alfred Needler and Teleost on the Scotian Shelf during the summer of 2005. Can. Tech. Rep. Fish. Aquat. Sci. 2824: iv+25 p.

Fraser, H.M., Greenstreet, S.P.R., Fryer, R.J., and Piet, G.J. 2008. Mapping spatial variation in demersal fish species diversity and composition in the North Sea: accounting for species- and size-related catchability in survey trawls. ICES J. Mar. Sci. 65: 531-538.

Fraser, H.M., Greenstreet, S.P.R., and Piet, G.J. 2007. Taking account of catchability in groundfish survey trawls: implications for estimating demersal fish biomass. ICES J. Mar. Sci. 64: 1800-1819.

Fung, T., Farnsworth, K.D., Reid, D.G., and Rossberg, A.G. 2012. Recent data suggest no further recovery in North Sea Large Fish Indicator. ICES J. Mar. Sci. 69: 235-239.

Gascuel, D., Bozec, Y.-M., Chassot, E., Colomb, A., and Laurans, M. 2005. The trophic spectrum: theory and application as an ecosystem indicator. ICES J. Mar. Sci. 62: 443-452.

Golani, D. 1998. Distribution of Lessepsian migrant fish in the Mediterranean. Italian Journal of Zoology 65 (Suppl. 1): 95-99.

Gorelick, R. 2011. Commentary: Do we have a consistent terminology for species diversity? The fallacy of true diversity. Oecologia 167: 885-888.

Gosselin, F. 2001. Lorenz partial order: the best known logical framework to define evenness indices. Community Ecology 2: 197-207.

Gotelli, N.J., and Colwell, R.K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4: 379-391.

Gotelli, N.J., and Colwell, R.K. 2011. Estimating species richness. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 39-54.

Gray, J.S. 2000. The measurement of marine species diversity, with an application to the benthic fauna of the Norwegian continental shelf. J. Exp. Mar. Biol. Ecol. 250: 23-49.

Gray, J.S., Clarke, K.R., Warwick, R.M., and Hobbs, G. 1990. Detection of initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North Sea. Mar. Ecol. Prog. Ser. 66: 285-299.

Green, R., and Chapman, P.M. 2011. The problem with indices. Mar. Poll. Bull. 62: 1377-1380.

Greene, C.H., and Pershing, A.J. 2003. The flip-side of the North Atlantic Oscillation and modal shifts in slope-water circulation patterns. Limnol. & Oceanogr. 48: 319-322.

Greenstreet, S.P.R. 2008. Biodiversity of North Sea fish: Why do the politicians care but marine scientists appear oblivious to this issue? ICES J. Mar. Sci. 65: 1515-1519.

Greenstreet, S.P.R., Fraser, H.M., Rogers, S.I., Trenkel, V.M., Simpson, S.D., and Pinnegar, J.K. 2012a. Redundancy in metrics describing the composition, structure, and functioning of the North Sea demersal fish community. ICES J. Mar. Sci. 69: 8-22.

Greenstreet, S.P.R., and Hall, S.J. 1996. Fishing and groundfish assemblage structure in the north-western North Sea: an analysis of long-term and spatial trends. J. Animal. Ecol. 65: 577-598.

Greenstreet, S.P.R., and Piet, G.J. 2008. Assessing the sampling effort required to estimate α species diversity in the groundfish assemblages of the North Sea. Mar. Ecol. Prog. Ser. 364: 181-197.

Greenstreet, S.P.R., and Rogers, S.I. 2006. Indicators of the health of the North Sea fish community: identifying reference levels for an ecosystem approach to management. ICES J. Mar. Sci. 63: 573-593.

Greenstreet, S.P.R., Rogers, S.I., Rice, J.C., Piet, G.J., Guirey, E.J., Fraser, H.M., and Fryer, R.J. 2011. Development of the EcoQO for the North Sea fish community. ICES J. Mar. Sci. 68: 1-11.

Greenstreet, S.P.R., Rogers, S.I., Rice, J.C., Piet, G.J., Guirey, E.J., Fraser, H.M., and Fryer, R.J. 2012b. A reassessment of trends in the North Sea Large Fish Indicator and a re-evaluation of earlier conclusions. ICES J. Mar. Sci. 69: 343-345.

Guy, C.S., Neumann, R.M. and Willis, D.W. 2006. New terminology for Proportional Stock Density (PSD) and Relative Stock Density (RSD): Proportional Size Structure (PSS). Fisheries 31(2): 86-87.

Hall, S.J., and Greenstreet, S.P. 1998. Taxonomic distinctness and diversity measures: responses in marine fish communities. Mar. Ecol. Prog. Ser. 166: 227-229.

Halliday, R.G., Van Guelpen, L., and Themelis, D.E. 2012. Demersal fish fauna of the continental slope off Nova Scotia, Canada, based on exploratory bottom trawl surveys in 1994–95. J. Northw. Atl. Fish. Sci. 44: 41-60.

Heink, U., and Kowarik, I. 2010. What criteria should be used to select biodiversity indicators? Biodivers. Conserv. 19: 3769-3797.

Heino, J. 2010. Are indicator groups and cross-taxon congruence useful for predicting biodiversity in aquatic ecosystems? Ecological Indicators 10: 112-117.

Heino, J., Paavola, R., Virtanen, R., and Muotka, T. 2005. Searching for biodiversity indicators in running waters: do bryophytes, macroinvertebrates, and fish show congruent diversity patterns? Biodivers. Conserv. 14: 415-428.

Heino, J., Tolonen, K.T., Kotanen, J., and Paasivirta, L. 2009. Indicator groups and congruence of assemblage similarity, species richness and environmental relationships in littoral macroinvertebrates. Biodivers. Conserv. 18: 3085-3098.

Heip, C. 1974. A new index measuring evenness. J. Mar. Biol. Asso. UK 54: 555-557.

Heip, C.H.R., Herman, P.M.J., and Soetaert, K. 1998. Indices of diversity and evenness. Océanis 24(4): 61-87.

Heller, R., and Siegismund, H. 2009. Relationship between three measures of genetic differentiation G_{ST}, D_{EST}, and G'_{ST}: how wrong have we been? Molecular Ecology 18: 2080-2083.

Heltshe, J.F., and Forrester, N.E. 1983. Estimating diversity using quadrat sampling. Biometrics 39: 1073-1076.

Heltshe, J.F., and Forrester, N.E. 1985. Statistical evaluation of the jackknife estimate of diversity when using quadrat samples. Ecology 66: 107-111.

Henderson, P.A. 2007. Discrete and continuous change in the fish community of the Bristol Channel in response to climate change. J. Mar. Biol. Asso. UK 87: 589-598.

Hering, D., Borja, A., Carvalho, L. and Feld, C.K. 2013. Assessment and recovery of European water bodies: key messages from the WISER project. Hydrobiologia 704: 1-9.

Heslenfeld, P., and Enserink, E.L. 2008. OSPAR Ecological Quality Objectives: the utility of health indicators for the North Sea. ICES J. Mar. Sci. 65: 1392-1397.

Hill, M.O. 1973. Diversity and evenness: a unifying notation and its consequences. Ecology 54: 427-431.

Hillebrand, H., and Mattiessen, B. 2009. Biodiversity in a complex world: consolidation and progress in functional biodiversity research. Ecology Letters 12: 1-15.

Hoffmann, S., and Hoffman, A. 2008. Is there a "true" diversity? Ecological Economics 65: 213-215.

Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A.J., Vandermeer, J., and Wardle, D.A. 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecological Monographs 75: 3-35.

Horsman, T.L., Serdynska, A., Zwanenburg, K.C.T., and Shackell, N.L. 2011. Report on the Marine Protected Area Network Analysis in the Maritimes Region, Canada. Can. Tech. Rep. Fish. Aquat. Sci. 2917: xi + 188 p.

Hubbell, S.P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Monographs in Population Biology 32. Princeton U.P., Princeton, N.J.: viv+377 p.

Hurlbert, S.H. 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52: 577-586.

ICES. 2005. Ecosystem effects of fishing: impacts, metrics, and management strategies. ICES Coop. Res. Rep, 272: 177 p.

ICES. 2009. Report of the ICES-NAFO Working Group on Deep-water Ecology (WGDEC), 9-13 March 2009. ICES CM 2009/ACOM:23. 94 p.

ICES. 2010. Report of the Working Group on Biodiversity (WGBIODIV), 22-26 February 2010, Lisbon, Portugal. ICES CM 2010/SSGEF:06. 97 p.

ICES. 2011a. Report of the Working Group on the Ecosystem Effects of Fishing Activities (WGECO), 13-20 April 2011, Copenhagen, Denmark. ICES CM 2011/ACOM:24. 166 p.

ICES. 2011b. Report of the Working Group on Biodiversity Science (WGBIODIV), 21-25 February 2011, ICES Headquarters, Copenhagen, Denmark. ICES CM 2011/SSGEF:02. 94 p.

ICES. 2011c. Report of the Workshop on Marine Biodiversity (WKMARBIO): furthering ICES engagement in biodiversity issues, 9-11 February 2011, Copenhagen, Denmark. ICES CM 2011/SSGSUE:02. 28 p.

ICES. 2012a. Report of the Working Group the Ecosystem Effects of Fishing Activities (WGECO), 11-18 April 2012, Copenhagen, Denmark. ICES CM 2012/ACOM:26. 192 p.

ICES. 2012b. Report of the Working Group on Biodiversity Science (WGBIODIV), 30 January – 3 February 2010, Nantes, France. ICES CM 2012/SSGEF:02. 98 p.

Intergovernmental Oceanographic Commission (IOC). 2006. A handbook for measuring the progress and outcomes of integrated coastal and ocean management. IOC Manuals and Guides 46; ICAM Dossier 2. UNESCO, Paris, France.

Ives, A.R., and Carpenter, S.R. 2007. Stability and diversity of ecosystems. Science 317: 58-61.

Jennings, S., Lancaster, J., Woolmer, A., and Cotter, J. 1999. Distribution, diversity and abundance of epibenthic fauna in the North Sea. J. Mar. Biol. Asso. UK 79: 385-399.

Jensen, A.L. 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. Can. J. Fish. Aquat. Sci. 53: 820-822.

Johnson, D. 2008. Environmental indicators: their utility in meeting the OSPAR Convention's regulatory needs. ICES J. Mar. Sci. 65: 1387-1391.

Jost, L. 2006. Entropy and diversity. Oikos 113: 363-375.

Jost, L. 2007. Partitioning diversity into independent alpha and beta components. Ecology 88: 2427-2439.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. Molecular Ecology 17: 4015-4026.

Jost, L. 2009a. Mismeasuring biological diversity: Response to Hoffman and Hoffman (2008). Ecological Economics 68: 925-928.

Jost, L. 2009b. *D vs* G_{ST}: Response to Heller and Siegismund (2009) and Ryman and Leimar (2009). Molecular Ecology 18: 2088-2091.

Jost, L. 2010a. The relation between evenness and diversity. Diversity 2: 207-232.

Jost, L. 2010b. Independence of alpha and beta diversities. Ecology 91: 1969-1974.

Jost, L., Chao, A., and Chazdon, R.L. 2011. Compositional similarity and β (beta) diversity. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 66-84.

Jurasinski, G., and Koch, M. 2011. Commentary: Do we have a consistent terminology for species diversity? We are on the way. Oecologia 167: 893-902.

Jurasinski, G., Retzer, V., and Beierkuhnlein, C. 2009. Inventory, differentiation, and proportional diversity: a consistent terminology for quantifying species diversity. Oecologia 159: 15-26.

Karakassis, I. 1995. S_∞: a new method for calculating macrobenthic species richness. Mar. Ecol. Prog. Ser. 120: 299-303.

Kelso, J.R.M., Steedman, R.J., and Stoddart, S. 1996. Historical causes of change in Great Lakes fish stocks and the implications for ecosystem rehabilitation. Can. J. Fish. Aquat. Sci. 53(Suppl. 1): 10-19.

Kenchington, E.L. 2003. The effects of fishing on species and genetic diversity. *In* Responsible Fisheries in the Marine Ecosystem. Edited by M. Sinclair and G. Valdimarsson. FAO, Rome & CABI Publishing, Wallingford. pp. 235-253.

Kenchington, E.L., and Heino, M. 2002. Maintenance of genetic diversity: challenges for management of marine resources. ICES CM 2002/Y:13.

Kenchington, E.L., Heino, M., and Nielsen, E.E. 2003. Managing marine genetic diversity: time for action? ICES J. Mar. Sci. 60: 1172-1176.

Kenchington, E.L.R., Kenchington, T.J., Henry, L.-A., Fuller, S., and Gonzalez, P. 2007. Multi-decadal changes in the megabenthos of the Bay of Fundy: The effects of fishing. J. Sea Research 58: 220-240.

Kenchington, E.L.R., Siferd, T., and Lirette, C. 2012. Arctic marine biodiversity: Indicators for monitoring coral and sponge megafauna in the Eastern Arctic. CSAS Res. Doc. 2013/003.

Kenchington, T.J. 1984. Population structures and management of the redfishes (*Sebastes* spp.: Scorpaenidae) of the Scotian Shelf. Ph.D. thesis, Dalhousie University: 491 p.

Kenchington, T.J. 1996. Long-term stability and change in the commercial groundfish longline fishing grounds of the northwest Atlantic. Fisheries Research 25: 139-154.

Kenchington, T.J. 2010. Environmental monitoring of the Gully Marine Protected Area: a recommendation. CSAS Res. Doc. 2010/075.

Kenchington, T.J. 2013a. A monitoring framework for the St. Anns Bank Area of Interest. CSAS Res. Doc. 2013/117.

Kenchington, T.J. 2013b Natural mortality estimators for information-limited fisheries. Fish and Fisheries. DOI: 10.1111/faf.12027.

Kenchington, T.J., Best, M., Bourbonnais-Boyce, C., Clement, P., Cogswell, A., MacDonald, B., MacEachern, W.J., MacIsaac, K., MacNab, P., Paon, L., Reid, J., Roach, S., Shea, E., Themelis, D., and Kenchington, E.L.R. 2009. Methodology of the 2007 Survey of Meso- and Bathypelagic Micronekton of the Sable Gully: Cruise TEM768. Can. Tech. Rep. Fish. Aquat. Sci. 2853: vi+91p.

Kenchington, T.J., and R.G. Halliday. 1994. A survey of fishing practices in the Scotia-Fundy Region groundfish longline fisheries. Can. MS Rep. Fish. Aquat. Sci. 2225: 642p.

Kershner, J., Samhouri, J.F., James, C.A., and Levin, P.S. 2011. Selecting indicator portfolios for marine species and food webs: a Puget Sound case study. PLoS ONE 6(10): e25248. doi: 10.1371/journal.pone.0025248.

Laugen, A.T., Engelhard, G.H., Whitlock, R., Arlinghaus, R., Dankel, D.J., Dunlop, E.S., Eikset, A.M., Emberg, K., Jørgensen, C., Matsumura, S., Nusslé, S., Urbach, D., Baulier, L., Boukal, D.S., Ernande, B., Johnson, F.D., Mollet, F., Pardoe, H., Therkildsen, N.O., Uusi-Heikkilä, S., Vainikka, A., Heino, M., Rijnsdorp, A.D., and Dieckmann, U. 2012. Evolutionary impact assessment: accounting for evolutionary consequences of fishing in an ecosystem approach to fisheries management. Fish and Fisheries doi: 10.1111/faf.12007.

Leinster, T., and Cobbold, C.A. 2012. Measuring diversity: the importance of species similarity. Ecology 93: 477-489.

Lengyel, N.L., Collie, J.S., and Valentine, P.C. 2009. The invasive colonial ascidian *Didemnum vexillum* on Georges Bank – ecological effects and genetic identification. Aquatic Invasions 4: 143-152.

Leonard, D.R.P., Clarke, K.R., Somerfield, P.J., and Warwick, R.M. 2006. The application of an indicator based on taxonomic distinctness for UK marine biodiversity assessments. J. Environ. Management 78: 52-62.

Loseto, L., Wazny, T., Cleator, H, Ayles, B., Cobb, D., Harwood, L., Michel C., Nielsen O., Paulic, J., Postma, L., Ramlal, P., Reist, J., Richard, P., Ross, P.S., Solomon, S., Walkusz, W., Weilgart, L., and Williams, B. 2010. Information in support of indicator selection for monitoring the Tarium Niryutait Marine Protected Area (TNMPA). CSAS Res. Doc. 2010/094.

MacArthur, R.H. 1957. On the relative abundance of bird species. Proc. Nat. Acad. Sci. Wash. 43: 293-295. [Cited by Pielou 1977]

MacCall, A.D. 2011. The sardine–anchovy puzzle. *In* Shifting Baselines: The past and future of ocean fisheries. Edited by J.B.C. Jackson, K.E. Alexander and E. Sala. Island Press, Washington. pp. 47-57.

Magurran, A.E. 1988. Ecological Diversity and its Measurement. Princeton U.P., Princeton: x+179 p.

Magurran, A.E. 2004. Measuring Biological Diversity. Blackwell, Oxford: viii+256 p.

Magurran, A.E. 2011. Measuring biological diversity in time (and space). *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 85-94.

Magurran, A.E., and Henderson, P.A. 2011. Commonness and rarity. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 97-104.

Magurran, A.E., and McGill, B.J. 2011. Challenges and opportunities in the measurement and assessment of biological diversity. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 1-7.

Margalef, D.R. 1958. Information theory in ecology. General Systems 3: 36-71. [Cited by Pielou 1977]

Markle, D.F., Scott, W.B., and Kohler, A.C. 1980. New and rare records of Canadian fishes and the influence of hydrography on resident and nonresident Scotian Shelf ichthyofauna. Can. J. Fish. Aquat. Sci. 37: 49-65.

Marques, J.C., Sales, F., Patrício, J., Teixeira, H., and Neto, J.M. 2009. Ecological Indicators for Coastal and Estuarine Environmental Assessment: A User Guide. WIT Press, Southampton: xi+183 p.

Marsh, R., Petrie, B., Weidman, C.R., Dickson, R.R., Loder, J.W., Hannah, C.G., Frank, K., and Drinkwater, K. 1999. The 1882 tilefish kill – a cold event in shelf waters off the north-eastern United States? Fisheries Oceanography 8: 39-49.

Maurer, B.A., and McGill, B.J. 2011. Measurement of species diversity. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 55-65.

Mayer, P. 2006. Biodiversity – the appreciation of different thought styles and values helps to clarify the term. Restoration Ecology 14: 105-111.

McCallum, B.R., and Walsh, S.J. 1997. Groundfish survey trawls used at the Northwest Atlantic Fisheries Centre, 1971 to present. NAFO Sci. Coun. Studies 29: 93-104.

McCallum, B.R., and Walsh, S.J. 2001. Evaluating the success of the survey trawl standardization program at the Northwest Atlantic Fisheries Centre. NAFO SCR Doc. 01/26, Serial No. N4399.

McCann, K.S. 2000. The diversity–stability debate. Nature 405: 228-233.

McGill, B.J. 2010. Towards a unification of unified theories of biodiversity. Ecology Letters 13: 627-642.

McGill, B.J. 2011. Species abundance distributions. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 105-122.

McRae, L., Loh. J., Bubb, P.J., Baillie, J.E.M., Kapos, V., and Collen, B. 2008. The Living Planet Index - Guidance for National and Regional Use. UNEP–World Conservation Monitoring Centre, Cambridge: 12 p.

Meirmans, P.G., and Hedrick, P.W. 2011. Assessing population structure: F_{ST} and related measures. Molecular Ecological Resources 11: 5-18.

MERCINA Working Group. 2001. Oceanographic responses to climate in the northwest Atlantic. Oceanography 14: 76-82.

Monte-Luna, P. del, Lluch-Belda, D., Serviere-Zaragoza, E., Carmona, R., Reyes-Bonilla, H., Aurioles-Gamboa, D., Castro-Aguirre, J.L., Guzmán del Próo, S.A., Trujillo-Millán, O., and Brook, B.W. 2007. Marine extinctions revisited. Fish and Fisheries 8: 107-122.

Moreno, C.E., and Rodríguez, P. 2011. Commentary: Do we have a consistent terminology for species diversity: Back to basics and towards a unifying framework. Oecologia 167: 889-892.

Moreno, M., Vezzulli, L., Marin, V., Laconi, P., Albertelli, G., and Fabiano, M. 2008. The use of meiofauna diversity as an indicator of pollution in harbours. ICES J. Mar. Sci. 65: 1428-1435.

Muniz, P., Venturini, N., Pires-Vanin, A.M.S., Tommasi, L.R., and Borja, Á. 2005. Testing the applicability of a Marine Biotic Index (AMBI) to assessing the ecological quality of soft-bottom benthic communities in the South America Atlantic region. Mar. Poll. Bull. 50: 624-637.

Muxika, I., Borja, Á., and Bonne, W. 2005. The suitability of the marine biotic index (AMBI) to new impact sources along European coasts. Ecological Indicators 5: 19-31.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. 70(12), Part I: 3321-3323.

Nei, M., and Roychoudhury, A.K. 1974. Sampling variances of heterozygosity and genetic distance. Genetics 76: 379-390.

Nicholson, M.D., and Jennings, S. 2004. Testing candidate indicators to support ecosystem-based management: the power of monitoring surveys to detect temporal trends in fish community metrics. ICES J. Mar. Sci. 61: 35-42.

Noss, R.F. 1990. Indicators for monitoring biodiversity: a hierarchical approach. Conservation Biology 4: 355-364.

Nye, J.A., Bundy, A., Shackell, N., Friedland, K.D., and Link, J.S. 2010. Coherent trends in contiguous survey time-series of major ecological and commercial fish species in the Gulf of Maine ecosystem. ICES J. Mar. Sci. 67: 26-40.

Olsgard, F., Brattegard, T., and Holthe, T. 2003. Polychaetes as surrogates for marine biodiversity: lower taxonomic resolution and indicator groups. Biodiversity and Conservation 12: 1033-1049.

Olsgard, F. and Gray, J.S. 1995. A comprehensive analysis of the effects of offshore oil and gas exploration and production on the benthic communities of the Norwegian continental shelf. Mar. Ecol. Prog. Ser. 122: 277-306.

Pardal, M.A., Cardoso, P.G., Sousa, J.P., Marques, J.C., and Raffaelli, D. 2004. Assessing environmental quality: a novel approach. Mar. Ecol. Prog. Ser. 267: 1-8.

Pauly, D., Christensen, V., Dalsgaard, J., Froese, R., and Torres, F. 1998a. Fishing down marine food webs. Science 279: 860-863.

Pauly, D., Christensen, V., and Walters, C. 2000. Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fisheries. ICES J. Mar. Sci. 57: 697-706.

Pauly, D., Froese, R., and Christensen, V. 1998b. How pervasive is "fishing down marine food webs"? Science 282: 183.

Pauly, D., and Watson, R. 2005. Background and interpretation of the 'Marine Trophic Index' as a measure of biodiversity. Phil. Trans. Roy. Soc. B 360: 415-423.

Pearcy, W.G., and Schoener, A. 1987. Changes in the marine biota coincident with the 1982–1983 *El Niño* in the northeastern subarctic Pacific Ocean. J. Geophys. Res. 92(C13): 14417-14428.

Peet, R.K. 1974. The measurement of species diversity. Annu. Rev. Ecol. Syst. 5: 285-307.

Pielou, E.C. 1966a. Shannon's formula as a measure of specific diversity: Its use and misuse. Amer. Nat. 100: 463-465.

Pielou, E.C. 1966b. The measurement of diversity in different types of biological collections. J. Theoret. Biol. 13: 131-144.

Pielou, E.C. 1974. Population and Community Ecology: Principles and methods. Gordon and Breach, New York: viii+424 p.

Pielou, E.C. 1975. Ecological Diversity. Wiley, New York: viii+165 p.

Pielou, E.C. 1977. Mathematical Ecology. Wiley, New York: x+384 p.

Pielou, E.C. 1984. The Interpretation of Ecological Data: A Primer on Classification and Ordination. Wiley, New York: xi+263 p.

Piet, G.J., and Jennings, S. 2005. Response of potential fish community indicators to fishing. ICES J. Mar. Sci. 62: 214-225.

Pomeroy, R.S., Parks J.E., and Watson, L.M. 2004. How is your MPA doing? A guidebook of natural and social indicators for evaluating Marine Protected Area management effectiveness. IUCN, Gland, Switzerland and Cambridge, England. xv+215 p.

Por, F.D. 1971. One hundred years of Suez Canal – A century of Lessepsian migration: retrospect and viewpoints. Systematic Biology 20: 138-159.

Quayle, J.F., and Ramsay, L.R. 2005. Conservation status as a biodiversity trend indicator: Recommendations from a decade of listing species at risk in British Columbia. Conservation Biology 19: 1306-1311.

Rachor, E., Reiss, H., Degraer, S., Duineveld, G.C.A., Van Hoey, G., Lavaleye, M., Willems, W., and Rees, H.L. 2007. Structure, distribution, and characterizing species of North Sea macro-zoobenthos communities in 2000. *In* Structure and dynamics of the North Sea benthos. Edited by H.L. Rees, J.D. Eggleton, E. Rachor and E. Vanden Berghe. ICES Coop. Res. Rep. 288. pp. 46-59.

Rees, H.L., Eggleton, J.D., Rachor, E., and Vanden Berghe, E. (eds.) 2007. Structure and dynamics of the North Sea benthos. ICES Coop. Res. Rep. 288: iii+259p.

Rees, H.L., Hyland, J.L., Hylland, K., Mercer Clarke, C.S.L., Roff, J.C., and Ware, S. 2008. Environmental indicators: utility in meeting regulatory needs. An overview. ICES J. Mar. Sci. 65: 1381-1386.

Reid, P.C., Edwards, M., Beaugrand, G., Skogen, M., and Stevens, D. 2003. Periodic changes in the zooplankton of the North Sea during the twentieth century linked to oceanic inflow. Fisheries Oceanography 12: 260-269.

Rice, J.C. 2000. Evaluating fishery impacts using metrics of community structure. ICES J. Mar. Sci. 57: 682-688.

Rice, J.C. 2003. Environmental health inidcators. Ocean & Coastal Manage. 46: 235-259.

Rice, J.C., and Rochet, M.-J. 2005. A framework for selecting a suite of indicators for fisheries management. ICES J. Mar. Sci. 62: 516-527.

Ricotta, C. 2004. A recipe for unconventional evenness measures. Acta Biotheoretica 52: 95-104.

Ricotta, C. 2005. Through the jungle of biological diversity. Acta Biotheoretica 53: 29-38.

Ricotta, C., De Zuliani, E., Pacini, A., and Avena, G.C. 2001. On the mutual relatedness of evenness measures. Community Ecology 2: 51-56.

Robinson, G.A., and Hunt, H.G. 1986. Continuous Plankton Records: Annual fluctuations of the plankton in the western English Channel, 1958–83. J. Mar. Biol. Asso. UK 66: 791-802.

Rodda, G.H. 1993. How to lie with biodiversity. Conservation Biology 7: 959-960.

Roff, D.A. 1984. The evolution of life history parameters in teleosts. Can. J. Fish. Aquat. Sci. 41: 989-1000.

Rogers, S.I., and Greenaway, B. 2005. A UK perspective on the development of marine ecosystem indicators. Mar. Poll. Bull. 50: 9-19.

Rosenberg, R., Blomqvist, M., Nilsson, H.C., Cederwall, H., and Dimming, A. 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. Mar. Poll. Bull. 49: 728-739.

Rosenzweig, M.L. 1995. Species Diversity in Space and Time. Cambridge University Press, Cambridge. xxi+436 p.

Rosenzweig, M.L., Donoghue, J., Li, Y.M., and Yuan, C. 2011. Estimating species density. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 276-288.

Rosindell, J., Hubbell, S.P., and Etienne, R.S. 2011. The unified neutral theory of biodiversity and biogeography at age ten. Trends in Ecology and Evolution 26: 340-348.

Royal Society. 2003. Measuring Biodiversity for Conservation. Royal Society Policy Document 11/03: viii+56 p.

Ruzzante, D.E., Taggart, C.T., Cook, D., and Goddard, S. 1996. Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland: microsatellite DNA variation and antifreeze level. Can. J. Fish. Aquat. Sci. 53: 634-645.

Ruzzante, D.E., Wroblewski, J.S., Taggart, C.T., Smedbol, R.K., Cook, D., and Goddard, S.V. 2000. Bay-scale population structure in coastal Atlantic cod in Labrador and Newfoundland, Canada. J. Fish. Biol. 56: 431-447.

Ryman, N., and Leimar, O. 2009. G_{ST} is still a useful measure of genetic differentiation – a comment on Jost's *D*. Molecular Ecology 18: 2084-2087.

Salas, F., Patrício, J., Marcos, C., Pardal, M.A., Pérez-Ruzafa, A., and Marques, J.C. 2006. Are taxonomic distinctness measures compliant to other ecological indicators in assessing ecological status?. Mar. Poll. Bull. 52: 817-829.

Samadi, S., Schlacher, T., and de Forges, B.R. 2007. Seamount benthos. *In* Seamounts: Ecology, Fisheries & Conservation. Edited by T.J. Pitcher, T. Morato, P.J.B. Hart, M.R. Clark, N. Haggan and R.S. Santos. Blackwell, Oxford. pp. 119-140.

Sarkar, S. 2010. Diversity: a philosophical perspective. Diversity 2: 127-141.

Schratzberger, M., Warr, K., and Rogers, S.I. 2006. Patterns of nematode populations in the southwestern North Sea and their link to other components of the benthic fauna. J. Sea Res. 55: 113-127.

Schratzberger, M., Warr, K., and Rogers, S.I. 2007. Functional diversity of nematode communities in the southwestern North Sea. Mar. Environ. Res. 63: 368-389.

Scott, W.B., and Scott, M.G. 1988. Atlantic Fishes of Canada. Can. Bull. Fish. Aquat. Sci. 219: 731p.

Shackell, N.L., Fisher, J.A.D., Frank, K.T., and Lawton, P. 2012. Spatial scale of similarity as an indicator of metacommunity stability in exploited marine systems. Ecol. Applications 22: 336-348.

Shackell, N.L., and Frank, K.T. 2003. Marine fish diversity on the Scotian Shelf, Canada. Aquatic Conserv.: Mar. Freshw. Ecosyst. 13: 305-321.

Shephard, S., Reid, D.G., and Greenstreet, S.P.R. 2011. Interpreting the large fish indicator for the Celtic Sea. ICES J. Mar. Sci. 68: 1963-1972.

Sherwin, W.B., Jabot, F., Rush, R., and Rossettos. 2006. Measurement of biological information with applications from genes to landscapes. Molecular Ecology 15: 2857-2869.

Shin, Y.-J., Rochet, M.-J., Jennings, S., Field, J.G., and Gislason, H. 2005. Using size-based indicators to evaluate the ecosystem effects of fishing. ICES J. Mar. Sci. 62: 384-396.

Shin, Y.-J., and Shannon, L.J. 2010. Using indicators for evaluating, comparing, and communicating the ecological status of exploited marine ecosystems. 1. The IndiSeas project. ICES J. Mar. Sci. 67: 686-691.

Shin, Y.-J., Bundy, A., Shannon, L.J., Simiar, M., Coll, M., Fulton, E.A., Link, J.S., Jouffre, D., Ojaveer, H., Mackinson, S., Heymans, J.J., and Raid, T. 2010a. Can simple be useful and reliable? Using ecological indicators to represent and compare the states of marine ecosystems. ICES J. Mar. Sci. 67: 717-731.

Shin, Y.-J., Shannon, L.J., Bundy, A., Coll, M., Aydin, K., Bez, N., Blanchard, J.L., Borges, M.de F., Diallo, I., Diaz, E., Heymans, J.J., Hill, L., Johannesen, E., Jouffre, D., Kifani, S., Labrosse, P., Link, J.S., Mackinson, S., Masski, H., Möllmann, C., Neira, S., Ojaveer, H., Abdallahi, K.o.M., Perry, I., Thiao, D., Yemane, D., and Curry, P.M. 2010b. Using indicators for evaluating, comparing, and communicating the ecological status of exploited marine ecosystems. 2. Setting the scene. ICES J. Mar. Sci. 67: 692-716.

Smedbol, R.K., and Stephenson, R. 2001. The importance of managing within-species diversity in cod and herring fisheries of the north-western Atlantic. J. Fish Biol. 59 (Suppl. A): 109-128.

Smith, B., and Wilson, J.B. 1996. A consumer's guide to evenness indices. Oikos 76: 70-82.

Smith, S.H. 1968. Species succession and fishery exploitation in the Great lakes. J. Fish. Res. Bd Can. 25: 667-693.

Somerfield, P.J., Clark, K.R., Warwick, R.M., and Dulvy, N.K. 2008. Average functional distinctness as a measure of the composition of assemblages. ICES J. Mar. Sci. 65: 1462-1468.

Soulé, M.E., and Wilcox, B.A. (eds.) 1980. Conservation Biology: An evolutionary-ecological perspective. Sinauer Associates, Sunderland, MA. xv+395 p.

Spellerberg, I.F., and Fedori, P.J. 2003. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon-Wiener' index. Global Ecology & Biogeography 12: 177-179.

Stachowicz, J.J., Best, R.J., Bracken, M.E.S., and Graham, M.H. 2008. Complementarity in marine biodiversity manipulations: reconciling divergent evidence from field and mesocosm experiments. PNAS 105: 18842-18847.

Stachowicz, J.J., Bruno, J.F., and Duffy, J.E. 2007. Understanding the effects of marine biodiversity on communities and ecosystems. Ann. Rev. Ecol. Evol. Syst. 38: 739-766.

Swingland, I.R. 2001. Biodiversity, definition of. *In* Encyclopedia of Biodiversity, Volume 1. Academic Press. pp. 377-391.

Teixeira, H., Weisberg, S.B., Borja, A., Ranasinghe, J.A., Cadien, D.B., Velarde, R.G., Lovell, L.L., Pasko, D., Phillips, C.A., Montagne, D.E., Ritter, K.J., Salas, F., and Marques, J.C. 2012. Calibration and validation of the AZTI's Marine Biotic Index (AMBI) for southern California marine bays. Ecological Indicators 12: 84-95.

Tuomisto, H. 2010a. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. Ecography 33: 2-22.

Tuomisto, H. 2010b. A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. Ecography 33: 23-45.

Tuomisto, H. 2010c. A consistent terminology for quantifying species diversity? Yes, it does exist. Oecologia 164: 853-860.

Tuomisto, H. 2011. Commentary: Do we have a consistent terminology for species diversity? Yes, if we choose to use it. Oecologia 167: 903-911.

Tuomisto, H. 2012. An updated consumer's guide to evenness and related indices. Oikos 121: 1203-1218.

Turner, J. 1994. The quality of wildness: preservation, control and freedom. *In* Place of the Wild: A wildlands anthology. Edited by D.C. Burks. Island Press, Washington. pp. 175-189.

Turvey, S.T. 2009. In the shadow of the megafauna: prehistoric mammal and bird extinctions across the Holocene. *In* Holocene Extinctions. Edited by S.T. Turvey. Oxford U.P., Oxford. pp. 17-39.

van Wijk, S.J., Taylor, M.I., Creer, S., Dreyer, C., Rodrigues, F.M., Ramnarine, I.W., van Oosterhout, C., and Carvalho, G.R. 2013.

Experimental harvesting of fish populations drives genetically based shifts in body size and maturation. Front. Ecol. Environ. 11: 181-187.

Veech, J.A., and Crist, T.O. 2010a. Diversity partitioning without statistical independence of alpha and beta. Ecology 91: 1964-1969.

Veech, J.A., and Crist, T.O. 2010b. Toward a unified view of diversity partitioning. Ecology 91: 1988-1992.

Vellend, M., Cornwell, W.K., Magnuson-Ford, K., and Mooers, A.Ø. 2011. Measuring phylogenetic biodiversity. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 194-207.

Ware, D.M. 1995. A century and a half of change in the climate of the NE Pacific. Fisheries Oceanography 4: 267-277.

Warwick, R.M., and Clark, K.R. 1995. New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. Mar. Ecol. Prog. Ser. 129: 301-305.

Warwick, R.M., and Clark, K.R. 1998. Taxonomic distinctness and environmental assessment. J. App. Ecology 35: 532-543.

Washington, H.G. 1984. Diversity, biotic and similarity indices: A review with special relevance to aquatic ecosystems. Water Res. 18: 653-694.

Weiher, E. 2011. A primer of trait and functional diversity. *In* Biological Diversity: Frontiers in measurement and assessment. Edited by A.E. Magurran and B.J. McGill. Oxford University Press, Oxford. pp. 175-193.

Weisberg, S.B., Ranasinghe, J.A., Dauer, D.M., Schaffner, L.C., Diaz, R.J., and Frithsen, J.B. 1997: An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake bay. Estuaries 20: 149-158.

Whitfield, P.E., Gardner, T., Vives, S.P., Gilligan, M.R., Coutenay, W.R., Ray, G.C., and Hare, J.A. 2002. Biological invasion of the Indo-Pacific lionfish *Pterois volitans* along the Atlantic coast of North America. Mar. Ecol. Prog. Ser. 235: 289-297.

Whitlock, M.C. 2011. G'_{ST} and D do not replace F_{ST}. Molecular Ecology 20: 1083-1091.

Wiersma, Y.F., and Urban, D.L. 2005. Beta diversity and nature reserve system design in the Yukon, Canada. Conservation Biology 19: 1262-1272.

Wilson, E.O. (ed.) 1988. Biodiversity. National Academy Press, Washington: xiii+521 p.

Xu, H., He, C., Xu, H., Huang, Y. and Sun, H. 2012. Application of taxonomic distinctness indices of littoral macroinvertebrate communities for assessing long-term variation in ecological quality status of intertidal ecosystems, northern China. Environ. Sci. Pollut. Res. 19: 3859-3867.

Xu, H., Jiang, Y., Al-Rasheid, K.A.S., Al-Farraj, S.A., and Song, W. 2011. Application of an indicator based on taxonomic relatedness of ciliated protozoan assemblages for marine environmental assessment. Environ. Sci. Pollut. Res. 18: 1213-1221.

Zahl, S. 1977. Jackknifing an index of diversity. Ecology 58: 907-913.

Zettler, M.L., Schiedek, D., and Bobertz, B. 2007. Benthic biodiversity indices versus salinity gradient in the southern Baltic Sea. Mar. Poll. Bull. 55: 258-270.

APPENDIX I

GLOSSARY AND LIST OF SYMBOLS

This glossary provides definitions for each term that is used in the main text with a specific meaning that is either unique to this report or at least less than universally accepted among specialists. Some terms and abbreviations that are not broadly familiar are included, even though the meanings they are here given are standard. This glossary is not, however, an explanation of all of the terms used in the report, nor a dictionary of the terminology of diversity theory.

Where it is necessary to use one of the defined terms in the text with some other meaning than the one given here, the word or expression is enclosed in quotation marks.

The list below also includes the abbreviations and mathematical symbols used in the text. In general, we have used one standard set of symbols throughout, altering those in original publications as necessary. However, some symbols stand as well-recognized labels for particular indices and those we have perforce maintained in their familiar forms, even at the cost of giving multiple meanings to some symbols.

α, γ and ε Diversities: Inventory diversities measured at habitat, landscape and regional levels respectively. In practice, ε diversity is rarely invoked, while α and γ diversities are applied to any two levels in an ecological hierarchy that may be of interest. Where more than two levels are studied simultaneously, the γ diversity of the lowest-level pair is treated as the α' of the next pair and so on.

β and δ Diversities: Differentiation diversities measured between habitat and landscape and between landscape and regional levels respectively. In practice, δ diversity is rarely invoked. If more than two levels in an ecological hierarchy are studied simultaneously, the differentiation diversities between each pair of levels are treated as β, β' etc.

Λ^+: Variation in Taxonomic Distinctness of Clarke and Warwick (2001):

$$\Lambda^+ = \frac{\sum\sum_{i \neq j} w_{ij}}{S(S-1)} - {\Delta^+}^2$$

Δ: In the context of genetic diversity: Symbol proposed by Jost (2008) for the Hill Numbers, ${}^{q}D$, when they are applied to allele frequencies. The value of q is indicated using the format $\Delta_{(q=2)}$, for example, except for $q = 1$, which is indicated as $\Delta_{(q-1)}$.

Appears in the forms Δ_S, Δ_T and Δ_{ST}, where the subscripts have the meanings outlined below (see *S*, *T* and *ST*).

Δ: In the context of phylogenetic diversity: Taxonomic Diversity of Warwick and Clarke (1995):

$$\Delta = \frac{\sum\sum_{i<j} w_{ij} p_i p_j}{\sum\sum_{i<j} p_i p_j + \sum_i p_i (p_i - 1)/2} = \frac{\sum\sum_{i<j} w_{ij} p_i p_j}{N(N-1)/2}$$

Δ^*: Taxonomic Distinctness of Warwick and Clarke (1995):

$$\Delta^* = \frac{\sum\sum_{i<j} w_{ij} p_i p_j}{\sum\sum_{i<j} p_i p_j}$$

Δ^+: Average Taxonomic Distinctness of Clarke and Warwick (1998):

$$\Delta^+ = \frac{\sum\sum_{i<j} w_{ij}}{S(S-1)/2}$$

κ: Number of pooled samples (from a particular dataset) needed for a pooled value of *H* to reach its asymptote.

K: Total number of samples available in a particular dataset. Alternatively symbolized as *t*.

μ_q: Expected similarity among the members of any group of *q* individuals randomly chosen (with replacement) from the community.

ν: Number of samples.

AMBI: AZTI's Marine Biotic Index. M-AMBI is a multivariate extension of AMBI.

Biodiversity or **Biological Diversity**: As defined by the CBD and hence the variability of living organisms and biological systems, encompassing all forms of variability at all levels of organization. Ecological diversity is a sub-part of biodiversity, which also includes, among others, phylogenetic diversity, functional trait diversity, intra-specific genetic diversity and distributions of individual body sizes.

BIP: Biodiversity Indicators Partnership.

CBD: Convention on Biological Diversity, of 1992.

Chao Estimator: Estimator of a lower bound on asymptotic *S*. There are two forms:

$$Chao1 \cdots S_{est} = S_{obs} + {f_1^2}/{2f_2}$$

suited to abundance data from a random collection of individuals, and:

$$Chao2 \quad \cdots \quad S_{est} = S_{obs} + \left(1 - \frac{1}{t}\right) {Q_1^2}\big/{2Q_2}$$

suited to presence / absence data from randomly distributed collections.

Community: Any group of organisms within defined spatial and temporal boundaries. The term "community" is applied to the bounded unit for which ecological diversity is quantified.

Community Composition: The relative abundances of named species, in a sample or a pool of samples. Community composition can be contrasted with species abundance data and SADs, which ignore the identifications of the species.

CU: Compositional Unit, a hypothetical community containing an effective number of species, each unique to that CU, which number is numerically equal to the average α diversity (itself measured as ^qD) at the lower of two levels in an ecological hierarchy (e.g. habitats in a comparison of habitats within landscapes). CUs are the units of β diversity in Tuomisto's (2010a) re-definition of that concept.

D: In the context of ecological diversity: Simpson's index of concentration:

$$D = \sum_{i=1}^{S} p_i^{\,2}$$

D: In the context of genetic diversity: Jost's (2008) "true differentiation".

$$D = 1 / \Delta_{ST} = \Delta_S / \Delta_T$$

***D*, *D*$_{max}$, *D*$_{min}$**: In the context of evenness: Respectively, the measured value of ecological diversity *sensu stricto* for the community of interest (using any preferred index), the maximum value that that D could take, given S for that community, and the minimum value that it could take under the same constraint.

^qD: Hill Number and Effective Number of Species (see "true" diversity).
Jost's (2006) symbol ^qD is used in this report but the older symbology introduced by Hill (1973) can still be found in the literature. Thus, what are here denoted as ^0D, ^1D and ^2D can be found represented by N_0, N_1 and N_2 (e.g. Greenstreet & Hall 1996; Callaway *et al.* 2002; Piet & Jennings 2005, Schratzberger *et al.* 2006; Greenstreet *et al.* 2012a).

$^q\overline{D}(T)$: Index of phylogenetic diversity derived by Chao *et al.* (2010) from ^qD:

$$^q\overline{D}(T) = \left\{ \sum_{i \in B_T} \frac{L_i}{T} a_i^q \right\}^{1/(1-q)}$$

(See main text for explanation of unique symbols in that equation.)

${}^qD^Z(p)$: Index of phylogenetic diversity derived by Leinster and Cobbold (2012) from ${}^qD$:

$${}^qD^Z(p) = \left\{ \sum_{i=1}^{S} p_i (Zp)_i^{q-1} \right\}^{1/(1-q)}$$

(See main text for explanation of unique symbols in that equation.)

d_{ij}: In the context of phylogenetic diversity: The phylogenetic distance between the i^{th} and j^{th} species in a tree. See also w_{ij}.

Differentiation Diversity: The difference in inventory diversity between levels in an ecological hierarchy, such as the greater diversity in a landscape when compared to those of the individual habitat patches within it. Usually encountered as β diversity.

E': Heip's index of evenness, derived from EXP H':

$$\tilde{E}' = \frac{e^{H'} - 1}{S - 1}$$

Ecological Diversity: The "diversity" of quantitative ecology, epitomized by the Shannon index. Strictly, ecological diversity is a combination of richness and evenness and in that form is here labelled "ecological diversity *sensu stricto*". However, many authors treat richness and evenness as alternative aspects of diversity. Thus, the combination of the three concepts is here termed "ecological diversity *sensu lato*".

Some authors have adopted "ecological diversity" for the γ diversity amongst ecosystems, since they insist that that cannot be "biological diversity" because ecosystems have abiotic components (e.g. Swingland 2001). In our usage, however, "ecological diversity" is not the diversity of ecosystems but the diversity of quantitative ecology. Magurran (2004) used "ecological diversity" for what is here termed "biological diversity", while she limited the latter term and its synonym, "biodiversity", to meaning species richness and relative abundances – more or less what is here termed "ecological diversity *sensu lato*".

Effective Number of Species: The number of species in a hypothetical community which shares the same value of a diversity index as the community being evaluated and in which every species is equally abundant (See: "true" diversity).

Environmental Status: All-embracing term, encompassing every facet of the condition of ecosystems, including biodiversity but also *inter alia* abundances, ranges, conservation risks, productivities, mortality rates, stabilities, ecosystem services and anthropogenic drivers of any of those.

Evenness: A measure of the variability in relative abundances of the species (or other kinds) present in a community. Also the factor that causes ecological diversity *sensu stricto* to deviate from richness. Evenness is usually quantified in terms of the latter meaning.

EXP H': Exponential Shannon diversity index.

Exponential Shannon Index: An index of ecological diversity *sensu stricto* which equals the exponential of the Shannon index, H', i.e. $e^{H'}$. Equal to ^1D.

Extirpation: Loss or withdrawal of a species or other taxon from a defined area.

Extinction: See "Global Extinction", "Local Extinction" and "Extirpation".

Facet of biodiversity or environmental status: Any aspect of biodiversity or environmental status that is deemed to merit attention. Facets may be as broad as intra-specific genetic diversity *in toto* or as narrow as the size composition of one particular population. Any attempt to "measure biodiversity" can at best yield a measurement of one facet and some authors have suggested that "biodiversity" is only definable in the context of particular facets.

FIE: Fisheries Induced Evolution.

f_r: The number of species represented in a dataset by r specimens each.

F_{ST}: In the context of genetic diversity: Wright's measure of gene differentiation for biallelic systems.

G_{ST}: In the context of genetic diversity: Nei's (1973) gene differentiation.

$$G_{ST} = D_{ST} / H_T = (H_T - H_S) / H_T$$

Genetic Diversity: Intra-specific diversity of alleles or other genetic markers.

Global Extinction: Entire elimination of a species or other described taxon from the biosphere.

H: In the context of ecological diversity: Brillouin's index of ecological diversity:

$$H = \frac{1}{N} \log_b \frac{N!}{n_1! n_2! \ldots n_S!}$$

H_{max}, H_{min}: As D_{max} and D_{min} but with ecological diversity quantified by Brillouin's H specifically.

H: In the context of genetic diversity: Heterozygosity – which, in a randomly-mating population, is equal to Nei's gene diversity:

$$H = 1 - \sum_{i=1}^{k} p_i^2$$

Appears in the forms H_S and H_T, where the subscripts have the meanings outlined below (see S and T).

H_i: Value of Brillouin's index for the i^{th} sample.

H_P: Allen's Phylogenetic Entropy:

$$H_p = -\sum_i L_i a_i \log a_i$$

(See main text for explanation of unique symbols in that equation.)

H': Shannon's index of ecological diversity *sensu stricto*:

$$H' = -\sum p_i \log_b p_i$$

H'': Biased approximation to Shannon's index of ecological diversity, in which p_i is replaced by N_i / N.

Habitat Diversity: Diversity of habitats. As used here, the term is broader than simply ecological diversity measured at the level of habitat patches, though that would be one approach to its quantification.

Hill Numbers: Generic family of indices of ecological diversity, advanced by Hill (1973) and taking the form:

$${}^{q}D = \left(\sum_{i=1}^{S} p_i^{\,q} \right)^{1/(1-q)}$$

Index: A quantifiable measure that serves to reduce multiple variables to a scalar value, in the form of a quantitative tool that can be applied to many datasets. An "index" is usually considered to be something more than a mere summary statistic (e.g. a mean) but nevertheless something that can be defined by a simple equation. Shannon's *H'* is a classic example.

Indicator: Situation-specific measure which may, but need not, invoke some index.

Inventory Diversity: Ecological diversity measured at any level in an ecological hierarchy, such as a habitat patch or a landscape, and standing in contrast to differentiation diversity. Usually encountered as sample, α or γ diversity.

J: Index of evenness, derived from Brillouin's *H* and taking the deficient form D/D_{max}. Replaced by *V*.

J': Pielou's index of evenness, derived from Shannon's *H'*.

$$\tilde{J}' = \frac{\tilde{H}'}{\log_b S}$$

K: Parameter of the von Bertalanffy growth model.

k: Number of alleles at a particular locus.

L_∞: Asymptotic length parameter of the von Bertalanffy growth model.

LFI: Large Fish Indicator.

Local Extinction: Entire elimination from the biosphere of a semi-discrete population that has not been formally described as an infra-species or sub-species. Also called **Population Extinction**.

LPI: Living Planet Index.

M: Natural mortality rate.

Measure: NOUN: Any unit of measurement. Includes, but is broader than, "index" and "metric".

VERB: The process of developing a quantitative value of an index or other measure. In practice, it more often involves estimation than measurement *per se*.

Measurement: The result of the measurement process, meaning the quantitative value obtained.

Metric: Synonym of "Index".

MTI: Marine Trophic Index, sometimes called the Mean Trophic Level.

MTL: Mean Trophic Level, sometimes called the Marine Trophic Index.

N: Total number of specimens gathered by a survey.

N_i or n_i: Number of specimens of the i^{th} species (or other kind) gathered by a survey.

Open sea: Any marine area more than a few nautical miles from the nearest land.

p_i: Relative abundance of the i^{th} species (or other kind).

Phylogenetic Diversity: Adaptation of ecological diversity *sensu stricto* that weights the species based on the degree of their relatedness, the latter derived from a tree that is founded in the results of either molecular taxonomy or numerical taxonomy .

Q: The inter-quartile gradient of a cumulative distribution function.

Q: In the context of phylogenetic diversity, Rao's Quadratic Entropy:

$$Q = \sum_{ij} d_{ij} p_i p_j$$

Q_r: The number of species represented in *r* collections within a dataset.

q: "Order" of a Hill Number and of "true" diversity.

r: In the context of Hurlbert's *V*, the product of *S* and the non-integer residual of *N*/*S*.

Richness: The number of kinds, types or varieties in a sample or larger group. Most often encountered as species richness – the number of species in a sample, community or other unit.

S: Species richness. When used as a parameter in an equation, *S* can be the number of species (or other kinds) used in the calculation and not necessarily the community richness.
The true richness of a taxocene (even a taxocene filtered through some survey protocol) within a community can only be determined by finding the

asymptote of a Taxon Sampling Curve. It is therefore sometimes termed the "asymptotic *S*". That can be contrasted with alternative values derived by rarefaction, which are here termed "rarefied *S*" and which underestimate asymptotic *S*.

S: As a subscript in the context of genetic diversity: Indicates the mean of the values for individual populations. See also *ST*.

S_{est}: Estimated lower bound on asymptotic *S*, derived from one of the Chao estimators.

S_{obs}: Number of species (or other kinds) observed in a particular dataset. Can be smaller than asymptotic *S*.

SAD: See Species Abundance Distribution.

SBI: Size-Based Indicator.

Sample Diversity: Ecological diversity within a single sample, such as the catch from a single trawl set or the sample recovered by a single drop of a grab. For some purposes, it is appropriate to maintain the conceptual linkage to the survey protocols through which the data were derived while considering pooled collections of material from two or more samples. The diversity in such a pooled collection is sometimes termed "pooled sample diversity".

Selectivity: Collective term encompassing the consequences of all processes, other than those of random sampling, that cause the relative abundances of species in a collection to differ from those in the community from which the collection was drawn. Includes, but is not limited to, the consequences of processes arising from the sampling gear used (e.g. mesh selection), those arising from when, where and how that gear was used, plus those driven by post-collection sample processing.

Selectivity in this sense has been called "detectability" but is closer to the "partial recruitment" of fisheries terminology.

Species Abundance Data: Data on the relative abundances of species, in a sample or a pool of samples. The identifications of the species are disregarded.

Species Abundance Distribution (SAD): Mathematical distribution of the numbers of species represented by N_i specimens (in a sample, a community or some larger unit), plotted against N_i.

Species Density: Similar to species richness but the richness seen in an area of some selected extent or in the sample obtained through some selected amount of survey effort.

Species Diversity: Ecological diversity measured at the species level.

Species Richness: Richness measured at the species level. Some authors limit it to the richness in a sample of *N* individual organisms, *N* being any chosen number.

ST: As a subscript in the context of genetic diversity: Indicates a value related to the difference between overall values and those of individual populations.

T: As a subscript in the context of genetic diversity, indicates the value for pooled populations. See also *ST*.

T: In the context of phylogenetic diversity: The time since the root of a tree or since some node of interest within a tree.

t: Total number of samples available in a particular dataset. Alternatively symbolized as K.

t_0: Parameter of the von Bertalanffy growth model.

Taxocene: Combination of all of the individuals that are members of one taxon, usually taken to mean a taxon higher than species. Used as the bounded group of organisms for which diversity is quantified, though such groups are not always literally taxocenes.

Taxonomic Diversity: Similar to phylogenetic diversity but with relatedness based on classic Linnean taxonomy.

Temporal Community: A community which shares the same spatial bounds as at least one other, from which it is divided in time.

"True" Diversity: Ecological diversity represented by the Hill Numbers, ^qD, which are taken to be the effective number of species (of Order q). See Jost (2006).

V: Hurlbert's index of evenness derived from Brillouin's H:

$$V = \frac{H - H_{\min}}{H_{\max} - H_{\min}}$$

where:

$$H_{\max} = \frac{1}{N} \log_b \frac{N!}{(X!)^{S-r}(Y!)^r}$$

$$H_{\min} = \frac{1}{N} \log_b \frac{N!}{(N - S + 1)!}$$

w_{ij}: In the context of Taxonomic Diversity, the "distinctness weight" of the path linking the i^{th} and j^{th} species in a tree. See also d_{ij}.

X: In the context of Hurlbert's V, the integer part of N/S.

Y: In the context of Hurlbert's V: X+1.

Z: Total mortality rate.

APPENDIX II

DETERMINING SPECIES DIVERSITY FROM AN EXISTING DATABASE

This Appendix draws on the material in the main text to provide a guide for practical application. Users should be familiar with the content of the rest of the report but section references are provided here to facilitate checking of details.

1: Determine the purpose of measuring ecological diversity and the questions to be answered using the measured values. The better the definition of why the measurements are wanted, the more likely that the correct choices of how to derive them will be made.

2: Develop an experimental design to answer the questions set under Step 1. The design may be as simple as a time series but could be much more complex (see Section 4.5).

3: Determine the unit under study:

    a: Select the database (see Section 4.1.4),

    b: Define the spatial boundaries of the community or communities (see Section 4.1.1),

    c: Define the temporal boundaries of the community or communities (see Sections 4.1.1 & 4.1.3), and

    d: Define the taxocene (see Section 4.1.2).

    All of those choices must be mutually compatible and the combination of them must address the purpose set under Step 1, through the experimental design selected under Step 2.

    a: The database must, at a minimum, include data on the relative abundances of species, gathered using a survey protocol sufficiently standardized that any variations in survey selectivities do not materially affect the measured values of ecological diversity. Since most of the Department's databases were gathered for purposes other than diversity quantification, it is the analyst's responsibility to ensure that the standardization was adequate for the new purpose and to demonstrate that adequacy (see Section 4.1.4).

    b: Examination of temporal trends needs only one spatial community, with boundaries that may be arbitrarily chosen, though it may be more informative to divide the data from the chosen database among two or more communities if there is reason to expect spatial variation in the temporal trends.

Examination of spatial variation needs at least two spatial communities and often as many as the available data can provide richness estimates for. Unless the boundaries of those areas can be drawn around well-defined biological units, arbitrary boundaries will be required. They must be selected such that the purpose set under Step 1 can be addressed and such that answers to the questions then posed will not be founded on artifacts of community definition (see Section 4.1.1). It is often advantageous to define the communities to be compared so that they cover equal areas, avoiding some complexities of the non-linear relationship between ecological diversity and community extent (see Section 4.2.1). However, equal-area communities will tend to show species diversity varying with the diversity of habitats per unit area, which may create misleading artifacts, depending on the purpose set under Step 1 (see Section 4.1.1).

The full extent of the chosen community or communities must fall within the spatial limits for which the chosen database can provide data unaffected by material variations in survey selectivities.

c: Ideally, examination of spatial variation should use data gathered over one brief period of time, either a continuous period (if the data are adequate to represent species abundances averaged around diel, tidal and seasonal cycles) or else a discontinuous set of periods (e.g. when the tide is below its mean level during daylight) within some longer duration (e.g. summer 2012). It will, however, often be necessary to pool data over longer durations. Whatever temporal bounds may be chosen, their combination with the dataset, taxocene and spatial community boundaries must lead to answers that address the purpose set under Step 1, without generating artifacts that substantially distort the measured values of ecological diversity. It is the analyst's responsibility to ensure that such artifacts are avoided and to demonstrate that they have been (see Section 4.1.3).

Examination of temporal trends requires that at least two periods (which can be considered as temporally-bounded communities: see Section 4.1.1) be defined, so that the ecological diversities at those times can be compared, whether each period is continuous or discontinuous. Often, it will be necessary to define as many separate periods as the available data can support richness estimates for. It is essential that the temporal distribution of data collection, within each period (or "temporal community"), is similarly distributed across various temporal cycles and non-cyclical patterns of change, so that the measured diversities are comparable (unless the purpose of the comparison is to examine differences across some cyclical or other changes, such as those between *El Niño* and *La Niña* periods). It is highly desirable that the periods being compared are of equal duration, since longer periods are expected to show higher diversity, all else being equal. It is the analyst's responsibility to ensure comparability of the periods selected for comparisons and to demonstrate that they are comparable, as well as to ensure that whatever comparisons may be drawn address the purpose set under Step 1.

The full temporal extent of the chosen community or communities must fall within the limits for which the chosen database can provide data unaffected by material variations in survey selectivities.

d: The taxocene must be one adequately represented in the chosen database (after limitation by the spatial and temporal boundaries of the chosen community), the adequacy being partly an issue of sufficient numbers of species and numbers of individuals for meaningful measurement of ecological diversity but also a matter of ensuring that sorting of catches and their identification to species has been sufficient (and particularly sufficiently standardized) throughout the collection of the data, such that no artifacts have been introduced which might substantially affect measurement of diversity. It is also important that the taxocene be one whose ecological diversity is relevant to the purpose set under Step 1. It is the analyst's responsibility to ensure that those requirements are met and to demonstrate that they have been.

It is recommended that the chosen taxocene truly include the members of a single taxon or at least those of a group of a few taxa (e.g. teleosts plus chondrichthyans), unless limited to the members of a single trophic level or trophic guild within a taxon. If it is deemed essential to define a "pseudo-taxocene", cutting across taxonomic groups, it is the analyst's responsibility to ensure that the resulting artifacts do not influence the conclusions drawn and to demonstrate that they do not (see Section 4.1.2).

The taxocene should be clearly and explicitly defined, so that there is no doubt of which species and life-stages were included and which excluded (see Section 4.1.2).

If the available data will support determination of values of ecological diversity for more than one taxocene, those should be developed in parallel, not by pooling the taxocenes.

4: Select the measure of "commonness" and rarity to be used (see Section 4.1.5). Often, it will be abundance but it might alternatively be biomass, percentage cover or something else. All diversity values that will need to be compared in order to answer the questions set under Step 1 must be determined using the same measure.

5: Review the database and ensure that, for each recorded species in the taxocene and each sample (e.g. each trawl-survey set), there is a correct value of the measure selected under Step 4. Deficient data that cannot be corrected must be deleted before analysis.

Adequacy of a database for some other purpose (e.g. estimation of resource biomasses) is not evidence that it is adequate for biodiversity measurement. It is the analyst's responsibility to ensure and demonstrate that adequacy (see Section 4.1.4).

6: For each determination of diversity (one per community / taxocene combination) that is needed in order to answer the questions posed in Step 1 using the experimental design set under Step 2, prepare a Taxon Sampling Curve, using repeated random re-sampling of the available data (see Sections 4.2.1 & 4.2.3).

  a: If every Curve reaches its asymptote, extract values for asymptotic *S* and its confidence limits.

  b: If some or all Curves fall short of their asymptotes: For each such Curve:

    i) Randomly order the samples used in preparing the Curve,

    ii) Pool them into an expanding pool, first one sample, then two, then three ... and so on,

    iii) For each pool, calculate the value of the appropriate Chao estimator,

    iv) Plot the Chao values against the size of the pool, and

    v) If the resulting curve reaches an asymptote, extract that as asymptotic *S* (see Section 4.2.3).

  c: If asymptotic *S* cannot be determined for some communities, consider revising the experimental design developed under Step 2 to delete the requirement for the missing information, while still answering the question posed under Step 1. Failing that:

    i) Use rarefaction and the Taxon Sampling Curves to develop a rarefied *S* for each community, rarefying to a sample size that allows determination of *S* for every community.

    ii) Examine the Taxon Sampling Curves and determine whether they all approach their asymptotes sufficiently closely that the values of rarefied *S* at least preserve the rank order of the corresponding asymptotic *S*.

    If not, discard the values of rarefied *S*. In that case, species richness cannot be determined using the available data.

    Comparisons among communities must use either the asymptotic *S* for each community or else the rarefied *S* after rarefaction to the same sample size. Rarefied and asymptotic *S* cannot be validly compared. If rarefied *S* is used, the analyst should be aware of the complications surrounding that approach (see Section 4.2.1).

7: To determine ecological diversity *sensu stricto*:

  a: Randomly order the samples available for each required determination of diversity,

  b: Pool those samples into an expanding pool, first one sample, then two, then three ... and so on, c: For each pool, calculate the value of Brillouin's *H*,

d: Plot *H* against the size of the pool,

e: Repeat steps (a) to (d) with multiple different randomizations of the samples,

f: Determine whether the plot reaches its asymptote within the number of available samples (see Section 4.3).

If not, ecological diversity *sensu stricto* cannot be determined. Either the experimental design set under Step 2 must be revised or else more data must be collected.

8: If the asymptote is reached, determine asymptotic *H* and its variance, using Zahl's jackknife method (see Section 4.3.2). That asymptotic *H* is numerically equal to the community value of Shannon's H'.

9: Exponentiate H' (i.e. find $e^{H'}$) and the corresponding values of confidence limits. That $e^{H'}$ is the measure of ecological diversity *sensu stricto* for the community / taxocene combination.

10: If values of both $e^{H'}$ and asymptotic *S* are available, calculate the value of Heip's evenness index (see Section 4.4). If the only values of richness that are available are of rarefied *S*, evenness cannot be calculated without risk of seriously misleading results.

11: If the data were collected using standard units of survey effort or can be re-arranged into amounts taken per unit effort: For each spatial or temporal community, determine the overall "abundance" (in the units used when determining diversity, e.g. numerical abundance or biomass) of the members of the taxocene, per unit survey effort (see Section 5.2). The "abundance" measure can be proportionate to absolute abundance (e.g. as a catch per standard trawl set), provided that the relationship between the measure and what is measured is effectively constant across the communities of interest. It is variations in overall abundance across those communities, rather than the absolute values, that is of concern.

12: Examine the values of richness, ecological diversity *sensu stricto*, evenness and overall abundance, making whatever comparisons among communities that the experimental design requires. Draw inferences to answer the questions set under Step 1.